MOLECULAR CHARACTERIZATION OF A MEGAGAMETOPHYTE-SPECIFIC ENHANCER IN *ARABIDOPSIS THALIANA*

Wei Yang

Research School of Biological Sciences (RSBS)

The Australian National University

Canberra, Australia

and

The Center for the Application of Molecular Biology to International Agriculture (CAMBIA)

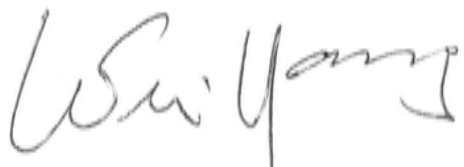
Canberra, Australia

A thesis submitted for the degree of Doctor of Philosophy of The Australian National University

June 2003

STATEMENT

I hereby declare that the results presented in this thesis are original work, except where acknowledged, conducted by myself in CAMBIA during the course of my enrolment as a PhD student at the Australian National University. No materials in this thesis have been previously submitted for an academic record at any other University.



Wei Yang

ACKNOWLEDGEMENTS

First of all, I would like to take this opportunity to express my deepest gratitude to my supervisor Dr. Richard A. Jefferson. It was him who opened the door to me for my Ph.D. study in the Australian National University when we first met in China National Rice Research Institute in Hangzhou, China and warmly hosted me to do my Ph.D. research in CAMBIA. Richard is not only an excellent scientific resource to me but also treats me as a friend. I would not have accomplished my Ph.D. program without his scientific guidance, powerful encouragement and sense of humor.

I would like to extend my deep gratitude to other members of my supervisory panel. Professor Murray Badger, as chair of the panel, always shows his gracious compassion and willingness to help. My successful *Arabidopsis* transformation was initiated with his kind help. Dr. Andrzej Kilian and Dr. Dean Price have always been helpful and I will always remember those rewarding discussions with them. Many good suggestions about the testing of enhancer elements have been from Dr. Paul Keese.

I also would like to thank Dr. Eric Huttner for critically reading my thesis. Thanks also goes to Professor Ueli Grossnikiaus of University of Zurich for providing me with the enhancer trap line ET253, Ming Luo of the Division of Plant Industry, CSIRO for teaching me the microscopic work on the observation of embryo and endosperm development, Prue Kell of RSBS, ANU for teaching me the *in planta* transformation of *Arabidopsis*.

During my graduate research in CAMBIA, I have been fortunate to be surrounded by people from many different countries who are all generous human beings. I thank those individuals, including Liz Deaves, Jorge Mayer, Carol Nottenburg, Chris Pratt, Selvameena Rajagopal, Leon Smith, Satya Nugroho, Xiqin Fu, Tuan Nguyen, Sri Koerniati, Thach Tran, Kaiman Peng, Sujin Patarapuwadol, Damian

Jaccoud, Peter Wenzl and Lijun Tian, for their friendship and for their advice on both scientific and interpersonal matters.

I also want to thank Dr. Toshio Murashige and Dr. Gary H. Toenniessen for recommending me to receive a Rockefeller Foundation Ph.D. fellowship. My special thanks goes to the Rockefeller Foundation for generous financial support for my graduate study.

Last, but not the least, I would like to thank my parents, my wife Ping Cao, and my sons Ziheng and Yaheng for their love, support and endurance, which has been my invaluable source of strength throughout my graduate study.

ABSTRACT

In *Arabidopsis thaliana*, large numbers of enhancer trap lines have been generated by several laboratories in the past several years and have been used in gene cloning and other studies of plant molecular biology, including the study of plant reproduction. This thesis presents results for the isolation, characterization and possible use of a megagametophyte-specific enhancer (MGSE) from a Ds-generated enhancer trap line ET253. TAIL-PCR and Inverse PCR were used to obtain the genomic region flanking the Ds insertion site in ET253. The Ds flanking sequence was then further analyzed by examining its capability to control *gus*A and GFP gene expression in the embryo sac when it was fused to a CaMV 35S minimal promoter and transformed back into wild type *Arabidopsis*.

A reliable enhancer testing system for testing DNA sequences with certain enhancer activity, especially those with tissue-specific activities in floral tissues was established in *Arabidopsis*. By using this system, a series of 5' and 3' deletion tests were carried out to determine the sequence responsible for the megagametophyte-specific activity. A 77 bp sequence with the MGSE activity was defined. Other data obtained in this study also showed that the MGSE is a unique cis-acting element in the *Arabidopsis* genome and it is quite conserved among different ecotypes. This MGSE is both position and orientation independent when tested in *Arabidopsis*. It is not associated with any nearby gene and therefore its endogenous role of the MGSE in the *Arabidopsis* genome remains a question.

The possibility of using the cloned MGSE for controlling specific gene expression in female gametophyte was also explored. By transactivation of the diphtheria toxin chain A (DTA) gene under the control of the MGSE in a Gal4-UAS transcription activation system, embryo specific cell ablation was achieved. This result demonstrated that the MGSE can be used as a tool to study and manipulate gene

expression in the female gametophyte and during early embryo development. The potential applications of the MGSE are also discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

UNDERSTANDING FEMALE GAMETOPHYTE DEVELOPMENT AND FUNCTION (An Introductory Literature Review)

1.1 Introduction

The normal plant life cycle is based on sexual reproduction, which involves the alteration between a diploid sporophytic generation and a haploid gametophytic generation. In flowering plants, the mature ovule consists of tissues from both a multicellular diploid sporophyte and a multicellular haploid female gametophyte (embryo sac) (Reiser and Fischer, 1993). It is the place in which sexual reproduction occurs and therefore allows the plant to exist with cyclic alteration of generations. In some circumstances, asexual reproduction (apomixis) can also occur when the sexual life cycle is short-circuited and seeds are set with all gene combinations identical to their maternal parent (Vielle-Calzada *et al.*, 1996). This trait could be of great importance for agricultural science, and in particular, for hybrid rice production (Jefferson, 1994; Grossniklaus *et al.*, 1998a).

Recently, the study of female gametophyte in plant has become increasingly attractive, partly driven by the hope for controllable apomixis. However, still very little is known so far about the genes and the pathways involved in the development of female gametophyte, although significant progress has been made in the genetic analyses of female gametophte development (Grossniklaus and Schneitz, 1998; Drews *et al.*, 1998). The reason for this is that the developmental stage of the ovule is brief, the tissue mass is very small and the mutants in this process are most likely to be lethal. The application of the insertional mutagenesis approaches using heterologous maize transposons or *Agrobacterium* mediated transfer DNA (T-DNA) insertions has greatly facilitated the work on the identification and isolation of novel plant genes that display a

mutant phenotype (Pereira, 2000). On the other hand, the development of gene trap and enhancer trap based on the maize transposable element Activator/Dissociation (Ac/Ds) or T-DNA insertion in *Arabidopsis* (Sundaresan *et al.*, 1995; Campisi *et al.*, 1999) provides valuable tools that permit identification of genes by their patterns of expression during development which can certainly avoid some disadvantages associated with mutagenesis approaches. Gene trap has been successfully used for the tagging of a PRL gene responsible for megagametophyte and embryo development (Springer *et al.*, 1995). Enhancer traps rely on a Ds element (DsE) or T-DNA carrying a *gusA* reporter gene under the control of a minimal promoter. If the DsE or T-DNA inserts in the proximity of a chromosomal gene, the result is the expression of the *gusA* gene in a specific temporal and spatial pattern under the control of the neighboring chromosomal enhancer or cis-regulatory element. This pattern could reflect the expression of the nearby gene controlled by the same regulatory element and thus allows the identification of genes according to their expression patterns rather than mutant phenotypes.

In this introductory chapter, I will first review the general biology of ovule and female gametophyte development, then focus on the studies towards understanding the molecular and genetic control of the female gametophyte development as well as the application of the enhancer trap system in tissue-specific gene and enhancer cloning.

1.2 Background: the importance of the study of female gametophyte in plant

As a major part of the female reproductive system and the progenitor of seed in higher plant, the ovule has attracted much interest to biologists since the last century (Schneitz *et al.*, 1995). The female gametophyte, which can also be called embryo sac or megagametophyte, is harbored in the ovule and plays a central role in the sexual

reproduction. It produces the egg cell and central cell, which give rise to embryo and endosperm respectively, following pollination (Reiser and Fischer, 1993). During the reproductive process, the female gametophyte is involved in many steps of interactions between the female and male gametophytic cells, including pollen tube guidance (Ray *et al.*, 1994; Higashiyama *et al.*, 1998), double fertilization (Russell, 1993) and the induction of seed development (Chaudhury *et al.*, 1998; Ohad *et al.*, 1999). A recent discovery that early embryo and endosperm development could mainly be under maternal control because the paternal genome showed a delayed activation (Vielle-Calzada *et al.*, 2000) certainly puts more weight on the role of the female gametophyte in early seed development.

A direct motivation for the study of the female reproductive system in higher plants has been closely related to apomixis, a natural process that allows clonal reproduction through seed and therefore keeps the progeny genetically identical to the mother plant. Over the last decade, molecular biologists have become much interested in understanding the mechanisms leading to modifications of the plant reproduction system capable of producing an apomictic phenotype and introducing apomixis into sexual crops by 'apomixis technology' (Koltunow *et al.*, 1995; van, I and van, I, 2000). If successfully implemented in the major crops, apomixis could bring more economic and social benefits than those of the green revolution (Grossniklaus *et al.*, 1998a). Apomictic processes share many events of sexual production but differ in the ways in embryo formation which give rise to fertile seeds. The apomictic embryo is derived solely from cells in the ovule rather than from the fusion of male and female gametes. As proposed by Nogler (Nogler, 1984) and further illustrated by Koltunow (Koltunow, 1993), three distinct apomictic processes, diplospory, apospory and adventitious embryony, are initiated at different times during ovule development. In diplospory, an unreduced embryo sac is formed when the megaspore mother cell switches from a sexual to an apomictic pathway. While in apospory, unreduced female gametophytes can be produced in one ovule with numerous cells differentiated from the nucellus.

These two processes are commonly referred to as gametophytic apomixis. In adventitious embryony, however, embryos can develop directly from somatic cells external to a sexual embryo sac and is therefore referred to as sporophytic apomixis. Although observations have been made extensively on these natural apomictic processes and some molecular analyses, such as markers linked to apomixis and comparative mapping, have been attempted, the genetic control of apomixis still remains unclear (Savidan, 2000). To ultimately control apomixis, female sexual reproductive function, especially female gametophyte development, has to be thoroughly investigated and better understood.

In the past several years, good progress has been made on the genetic analysis of female gametophyte development (Drews *et al.*, 1998; Yang and Sundaresan, 2000), mainly in *Arabidopsis*. However, molecular analysis of female gametophyte-specific gene expression has been hampered by the intractability of the tissue and progress has been slow. Some fundamental biological questions are still waiting to be clearly answered in the molecular level: What is the molecular basis of sporogenesis and megagametogenesis? What factors determine the commitment of special cells in the ovule to somatic, meiotic or apomictic development? How do gametes communicate during fertilization? To answer these questions, cell type-specific genes, promoters and enhancer elements regulating these processes have to be isolated and tested.

1.3 General biology of ovule and female gametophyte development: an overview

Ovule and embryo sac development in flowering plants has been well studied morphologically. Ovule morphologies are considerably diversified in different plants with various forms depending on the shape and position of the different ovary parts. In angiosperms, the most common type of ovule is the anatropous form, which shows a curvature of almost 180° during its development so that it becomes inverted and the

micropyle lies close to the funicle. While in gymnosperms, the orthotropous ovule, which shows funicle (funiculus), chalaza, embryo sac, and micropyle in a direct line, is the more common type (Bouman, 1984). Further distinction is made between campylotropous ovules with a slightly curved, kidney-shaped embryo sac, and amphitropous ovules with a more strongly curved, horseshoe-shaped embryo sac. The development of the embryo sac is also divided into the following types: monosporic (*Polygonum*-type and *Oenothera*-type), bisporic (*Allium*-type and *Endymion*-type) and tetrasporic (*Adoxa*-type etc), and over 15 different patterns have been described (Maheshwari, 1950; Willemse and van Went, 1984; Haig, 1990). However, the most common form is the *Polygonum*- type, which is found in about 70% observed species (Reiser and Fischer, 1993; Christensen *et al.*, 1998), and thought to be the ancestral type (Christensen *et al.*, 1998; Drews *et al.*, 1998).

Arabidopsis thaliana is a small ephemeral weed in the mustard or crucifer family (Brasicaceae or Cruciferae) (Price and Palmer, 1994). It possesses properties such as rapid growth, high fecundity (about 20,000 seeds per plant), short life cycle (6-8 weeks) and a small genome size (115-megabase pairs) and has therefore been chosen as a model system for studies of virtually all aspects of plant biology (Somerville and Meyerowitz, 1994). Like other angiosperms, *Arabidopsis thaliana* is heterosporous and has male and female gametophytes. Its male gametophyte (pollen or microgametophyte) develops within the anther and has a tricellular structure with two sperm cells encased in a vegetative cell. Its female gametophyte develops within the ovule, which has been characterized as either an anatropous or amphitropous type (Mansfield and Bowman, 1994). The *Arabidopsis* female gametophyte is of the popular *Polygonum* type consisting of seven cells: the egg, two synergids, a diploid central cell, and three antipodals (Webb and Gunning, 1990; Mansfield *et al.*, 1991). As a necessary prerequisite for a systematic genetic and molecular analysis of the whole process of the ovule development, a good classification scheme of ovule development stages using discrete, ovule-specific morphological markers has been established in

Arabidopsis (Webb and Bowman, 1994)(Schneitz *et al.*, 1995). Therefore, this overview will use *Arabidopsis* as a representative for the description of ovule and female gametophyte development.

1.3.1 Ovule development

Ovule is the direct precursor of seed and the development of a seed is a continuous process. Ovule development is commonly considered to be those processes occurring prior to fertilization. Further development after fertilization is referred as seed development. Ovule development in *Arabidopsis* takes place inside a typical *Brassicaceae*-type syncarpous gynoecium (two carpels united into a single structure) (Bowman *et al.*, 1999). It bears two rows of ovules in two locules separated by a septum with 20-30 ovules in each row. Ovules consist of a nucellus inside of which a megaspore mother cell and eventually an embryo sac (megagametophyte) develops. The nucellus is covered by two integuments originating from the chalaza region. The connection of the ovule to the placental region of a carpel is through a stalk termed funiculus. For ovule development to occur, several processes have to take place: initiation and outgrowth, specification of identity, pattern formation and morphogenesis. These steps can be regarded as independent processes (Schneitz *et al.*, 1995).

The morphological descriptions of ovule development in *Arabidopsis* have been well documented (Misra, 1962; Webb and Gunning, 1990; Mansfield *et al.*, 1991; Robinson-Beers *et al.*, 1992; Schneitz *et al.*, 1995). Figure 1.1 illustrates the main stages of ovule development in *Arabidopsis*. In the early phase of ovule development, the ovules originate through periclinal divisions in the sub-epidermal tissue of the placenta (inner ovary wall, referred to as L2 or L3 layer) to form the ovule primordia (nucellus) (Figure 1.1a). A single hypodermal cell (archesporial cell) at the distal end of the primordium differentiates into the megaspore mother cell (Figure 1.1b). At the same time, the inner integument is initiated by a series of cell divisions in the dermal (L1)

layer at a short distance behind the apex and the outer integument is initiated behind the inner integument shortly through a series of similar cell divisions in the dermal or subdermal layer. The divisions spread circumferentially and delimit the nucellus as the epical portion of the primordium. During this stage, the megaspore mother cell undergoes meiosis to produce a tetrad of haploid megaspores (Figure 1.1c). One of which survives and later gives rise to the mature embryo sac harboring the female reproductive cells (see detailed description in 1.2.2) as the body of the ovule enlarges and the funiculus elongates. During this process, the ovule exhibits the effects of asymmetric growth and both the nucellus and integuments curve forward. The outer integument gradually overtakes the inner integument and covers both the latter and the nucellus to produce the final amphitropous configuration (Figure 1.1d). The mature ovule has a small cleft (micropyle) at the distal end ready for the pollen tube to enter and conduct fertilization. The whole process from the initiation to mature ovule formation takes approximately 7 days.

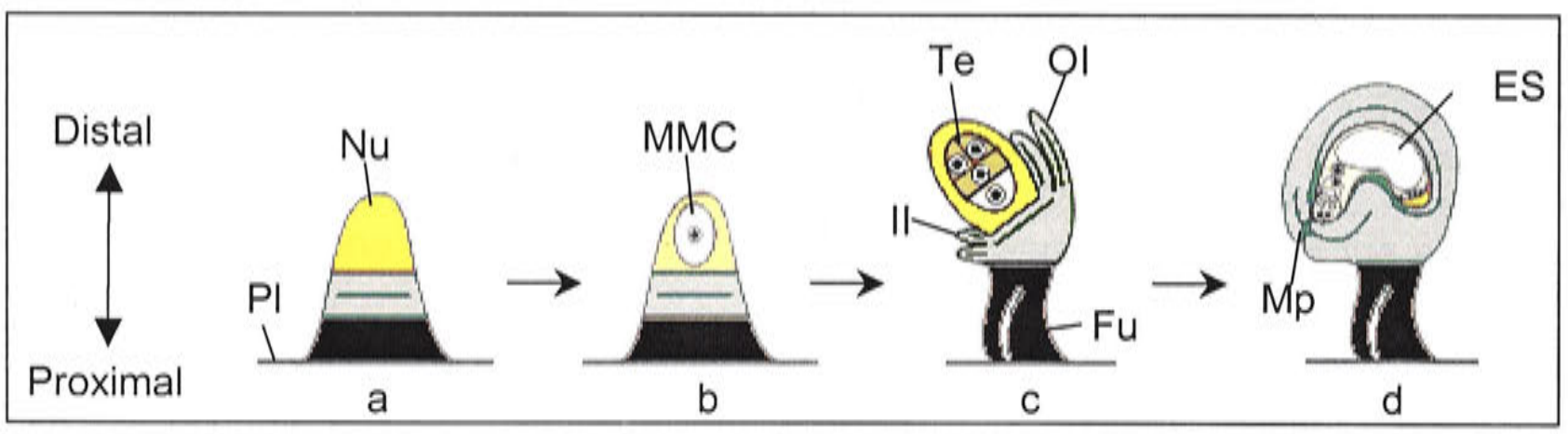


Figure 1.1 Cartoons mimicking the main stages of ovule development in *Arabidopsis thaliana* (modified from Grossniklaus and Schneitz, 1998).

Fu, funiculus; Il, inner integument; MES, mature embryo sac; MMC, megaspore mother cell; Mp, micropyle; Nu, nucellus; Ol, outer integument; Pl, placenta; Te, tetrad.

1.3.2 Embryo sac development

The *Arabidopsis* embryo sac is the typical monosporic *Polygonum*-type. The mature form of such embryo sac is shown in Figure 1.2. Generally, Polygonum-type embryo

sac development can be divided into two phases: megasporogenesis, which encompasses processes of megaspore mother cell differentiation, meiosis and the formation of a tetrad of megaspores; and megagametogenesis, which involves mitoses, nuclear positioning and cellularization to form a mature embryo sac (Maheshwari, 1950; Willemse and van Went, 1984; Reiser and Fischer, 1993; Drews *et al.*, 1998; Schneitz, 1999; Yang and Sundaresan, 2000).

The megagametogenesis starts with the formation of the megaspore mother cell (or megasporocyte). It is a large diploid cell located subdermally beneath the epidermal layer of the nucellus. During megagametogenesis, the megasporocyte undergoes meiosis to produce four haploid megaspores (tetrad) mainly in a linear arrangement. The chalazal-most megaspore survives to become functional, while the other three degenerate. This functional megaspore subsequently enters the megagametogenesis phase to form the embryo sac. During this process, the functional megaspore undergoes the first mitosis and the resulting two nuclei migrate to opposite poles to produce the two-nucleate embryo sac with a large central vacuole that was believed to play an important role in the nuclear positioning (Cass *et al.*, 1985). Each of the two nuclei then divides two more times, resulting in an eight-nucleate cell (embryo sac). Two polar nuclei, one from each pole, then migrate toward the center of the cell and fuse to form a diploid secondary endosperm nucleus. During polar nuclei migration, the eight-nucleate embryo sac cellularizes to produce seven cells: two synergid cells and an egg cell at the micropylar pole, three antipodal cells at the chalazal pole and a large central cell in the middle (Figure 1.2b). The polarity establishment is an important feature of female gametophyte development but the mechanism involved remains unknown (Drews *et al.*, 1998). This is the final structure of a mature embryo sac. An additional development after this is that the three antipodal cells degenerate before fertilization.

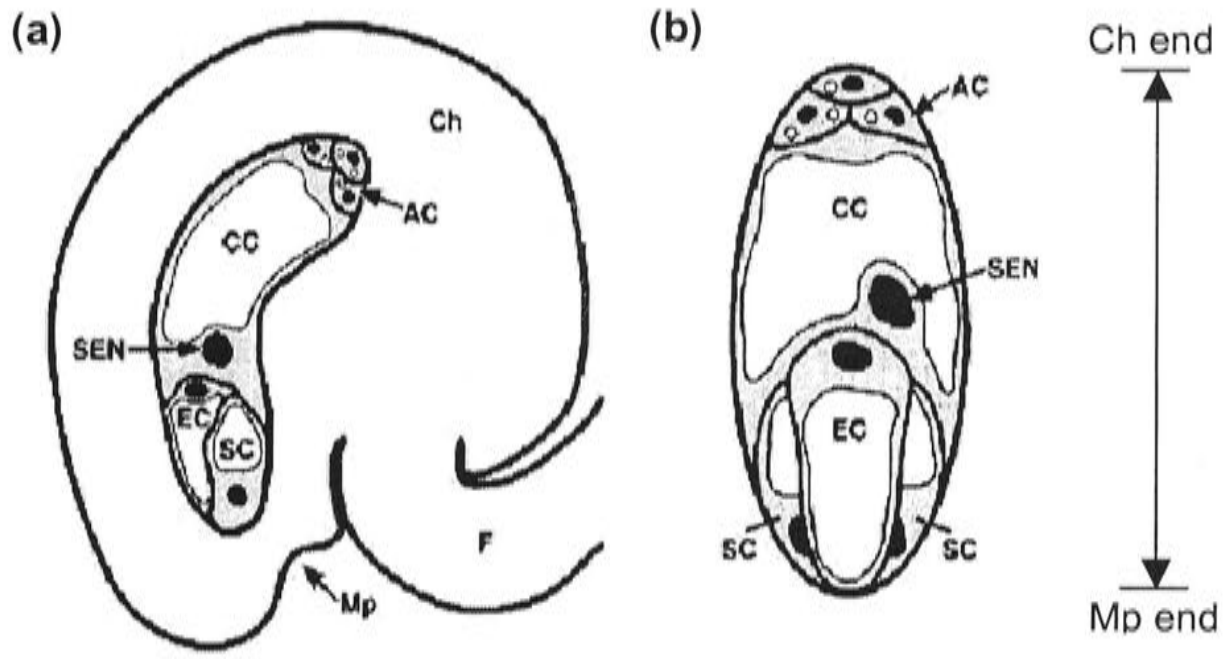


Figure 1.2 The Polygonum-type embryo sac of *Arabidopsis thaliana* (reproduced from Drews et al, 1998).

(a) Mature ovule of *Arabidopsis* with the view in longitudinal section showing the location of the embryo sac.

(b) Organization of the embryo sac. The gray, white and black areas represent cytoplasm, vacules and nuclei, respectively. The egg appratus, including the egg cell and synergids, is at the micropylar end. Three antipodal cells are at the chalazal end. A big central cell in the middle separates the egg appratus and the antipodal cells. AC, antipodal cells; CC, central cell; Ch, chalazal region; EC, egg cell; F, funiculus; Mp, micropyle; SC, synergid cell; SEN, secondary endosperm nucleus.

1.3.3 Molecular and genetic control of ovule development

As described in 1.2.1 and 1.2.2, the ontogeny of ovule in *Arabidopsis* is considered to be the result of four sequential processes: initiation, specification of identity, pattern formation and morphogenesis. These can be subdivided into three main parallel pathways of nucellar, chalazal and funicular development (Gasser *et al.*, 1998; Schneitz *et al.*, 1998). The molecular and genetic mechanisms controlling these processes are still poorly understood. However, progress has been made in genetic analysis of ovule development through mutations and some responsible genes have been cloned. Based on the available information, models of genetic regulation of ovule

development have been proposed which provide a good framework for future discoveries (Baker *et al.*, 1997; Gasser *et al.*, 1998; Schneitz *et al.*, 1998). Figure 1.3 outlines the current knowledge of the roles of genes involved in ovule development, although mutations in some of these genes revealed pleiotropic effects (e.g. some genes may play multiple roles in ovule and plant development). The following is a brief description of the possible roles of these genes in the corresponding pathways of ovule development.

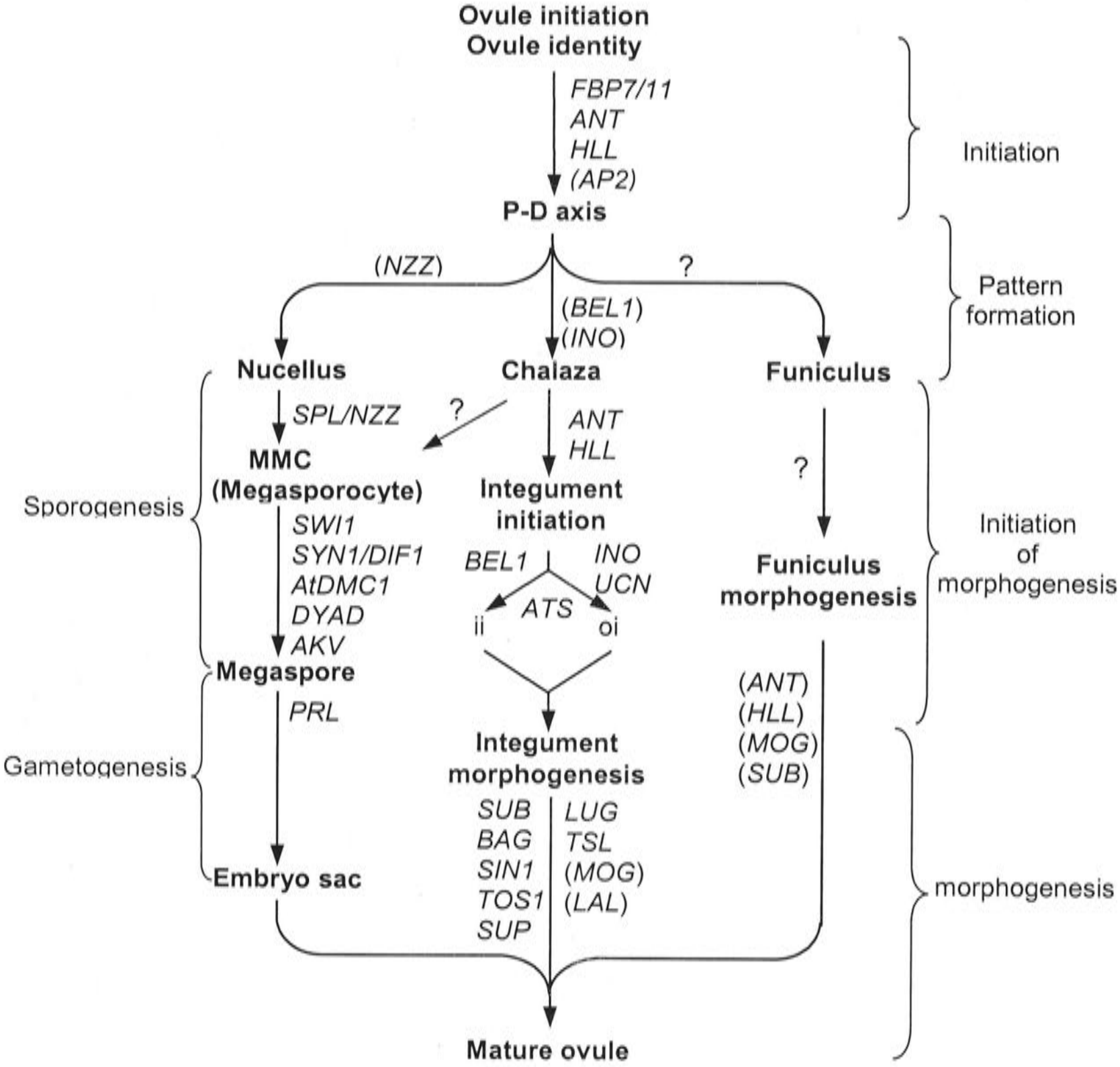


Figure 1.3 Genetic regulation of ovule development in *Arabidopsis*.

This model of regulation involved in ovule development is integrated from Schneitz *et al.* (1998), Gasser *et al.* (1998), and Yang and Sundaresan (2000).

Progress of ovule development is indicated from top to bottom by arrows. Gene designations are adjacent to the process they are believed to control or affect. Genes in parentheses indicate uncertainty about the gene function. Abbreviations: ii, inner integument; MMC, megaspore mother cell; oi, outer integument; P-D, proximal-distal. Full names of the genes are provided in the text.

1.3.3.1 Ovule initiation and identity

The molecular basis of specification and commitment from the placental tissue to the ovule fate is still not well understood and no gene with such a function has been isolated In *Arabidopsis*. However, two genes *FLORAL BINDING PROTEIN 7* (*FBP7*) and *FBP11* from *Petunia hybrida* have shown to be crucial for this process (Colombo *et al.*, 1995; Angenent *et al.*, 1995). Both genes encode putative transcription factors with a MADS domain. They are expressed in similar patterns with expression confined to the gynoecium. Initial expression is in the developing placenta but later is confined to the ovule primordia and then to the funiculus and emerging integuments (Angenent *et al.*, 1995). This confirms their role in ovule specification. Gasser *et al* (1998) hypothesized that there may be ortholog(s) of the *Petunia FBP7/ FBP11* genes in *Arabidopsis* that play similar role but it has not been proven yet.

AINTEGUMENTA (*ANT*) is a gene thought to be required repeatedly for floral development. It controls the initiation, growth and shape of floral organs in general, including ovules (Klucher *et al.*, 1996; Elliott *et al.*, 1996; Baker *et al.*, 1997; Schneitz *et al.*, 1997; Schneitz *et al.*, 1998). The *ANT* gene encodes a putative transcription factor homologous to the AP2 protein encoded by the homeotic gene *APETALA2* (*AP2*) (Jofuku *et al.*, 1994; Weigel, 1995). As for the effects on ovule development, ANT is necessary for normal initiation of integument primordia, for formation of two separate (inner and outer) primordia and for subsequent expansion of primordia, but may not play a direct role in megagametophyte development (Gasser *et al.*, 1998). Mutations in the *ANT* gene lead to the lack of both the inner and outer integuments (Klucher *et al.*,

1996; Elliott *et al.*, 1996). *AP2* may have a role in the promotion of ovule identity and initiation of female gametophyte development but may not be essential for ovule development (Gasser *et al.*, 1998; Byzova *et al.*, 1999). Another gene *HUELLENLOS* (*HLL*), which encodes a mitochondrial ribosomal protein (Skinner *et al.*, 2001), has overlapping function with *ANT* in regulating the outgrowth of the ovule primordium and therefore was thought to work in a partially redundant fashion (Schneitz, 1999).

1.3.3.2 Pattern formation

Pattern formation leads to the spatial arrangement of distinct regions with different cell fate within the ovule along the proximal-distal (P-D) or chalazal-micropylar axis, and the anterior-posterior (A-P) or adaxial-abaxial axis (Grossniklaus and Schneitz, 1998; Schneitz *et al.*, 1998). However, genes that directly function in the establishment of these patterns have not been identified. Some indirect evidence showed that two genes *BELL1* (*BEL1*) (Reiser *et al.*, 1995)and *INNER NO OUTER* (*INO*) (Baker *et al.*, 1997) may be involved in pattern formation of integuments. Another gene *NOZZLE* (*NZZ*) may perform an important role in the formation of the nucellus (Schneitz *et al.*, 1998). At present, no information is available for genes involved in the patterning of funiculus.

1.3.3.3 Morphogenesis

The first morphological change to occur in the ovule primordium is the emergence of the integuments. A number of genes have been identified that may regulate integument morphogenesis after pattern formation. They include *ANT*, *HLL*, *BEL1*, *INO*, *UNICORN* (*UCN*) (Schneitz *et al.*, 1997) and *ABERRANT TESTA SHAPE* (*ATS*) (Leon-Kloosterziel *et al.*, 1994). Besides their roles in ovule primordium outgrowth, *ANT* and *HLL* also have function in the initiation and growth of integuments. This means that these two genes are repeatedly required for cell proliferation process during ovule

development (Grossniklaus and Schneitz, 1998). Early integument development is also regulated by at least three genes *BEL1, INO* and *ATS*. *BEL1* and *INO*, beside their effects on integument specification, are also required for inner and outer integument initiation and growth, respectively. *ATS* is responsible for the formation of the boundary region between the two integuments (Leon-Kloosterziel *et al.*, 1994). On the other hand, the gene *UCN* may play a role as suppressor of outer integument initiation (Schneitz *et al.*, 1997).

Genes like *STRUBBELIG* (*SUB*), *BLASIG* (*BAG*) (Schneitz *et al.*, 1997), *SHORT INTEGUMENTS 1* (*SIN1*) (Robinson-Beers *et al.*, 1992; Ray *et al.*, 1996), *LEUNIG* (*LUG*) (Liu and Meyerowitz, 1995), TOUSLED (*TSL*) (Roe *et al.*, 1997), *TSO1* (Liu *et al.*, 1997; Hauser *et al.*, 1998; Song *et al.*, 2000) and *SUPERMAN* (*SUP*) (Gaiser *et al.*, 1995) may act after *ANT*, *HLL*, *BEL1* and *INO* to control the cell proliferation and shape of the integuments (Schneitz *et al.*, 1998). This is reflected by the fact that mutations in these genes normally do not lead to the absence of integuments, but only result in an altered morphology associated with later integumant development. For example, the *SUP* gene is thought to be involved in determination of the ovule symmetry to control the asymmetric growth of the outer integument (angenent et al, 1996). Two other genes, *MOLLIG* (*MOG*) and *LAELLI* (*LAL*), may also control integument development at certain stages but their exact roles remain to be understood (Schneitz *et al.*, 1997).

The genes involved in the funiculus morphogenesis are completely unknown. However, mutations in genes such as *ANT*, *HLL*, *MOG* and *SUB* also affect the vascular strand development, suggesting that they may be part of the funiculus development pathway (Schneitz *et al.*, 1998).

1.3.3.4 Megasporogenesis

Megasporogenesis starts with the differentiation of hypodermal cells of ovules to form archesporial cells. The differentiation of the archesporial cell then marks the beginning of the transition between the sporophytic and gametopytic phases of the plant cycle. Less is known about genes that are involved in the control and elaboration of these early steps of megasporogenesis. In *Arabidopsis*, only one gene, *SPOROCYTELESS* (*SPL*) or under the name of *NZZ*, is now known to be required for the initiation of sporogenesis (Yang *et al.*, 1999; Balasubramanian and Schneitz, 2000). In *spl/nzz* mutants, subepidermal cells of anther and ovule primordium are able to form archsporial cells but subsequently fail to form megasporocytes and microsporocytes. The *SPL/NZZ* gene encodes a nuclear protein with some similarities to MADS-box transcription factors and possibly functions in promoting the formation of both male and female sporocytes. However, *SPL/NZZ* might have other roles in addition to the development of sporocytes because it is also required for proximal-distal pattern formation and cell proliferation in ovule (1.3.3.2; (Balasubramanian and Schneitz, 2000). In maize, a locus *multiple archesporial cells1* (*mac1*), which plays an important role in the switch of the hypodermal cells from the vegetative to the meiotic (sporogenous) pathway in maize ovules, has been described (Sheridan *et al.*, 1996). However, no *mac1* equivalent has been identified in *Arabidopsis*.

Meiosis is a key step leading into gametogenesis. A number of genes and mutations affecting female meiosis have been reported and reviewed by Bhatt *et al.* (Bhatt *et al.*, 2001). It is thought that different genes may be employed during male and female meiosis in plants (Yang and Sundaresan, 2000). The fact that many mutations disrupting meiosis affect both sexes, however, implies that a common set of genes required for meiosis may be integrated into two overlapping but distinct developmental pathways leading to the formation of the male and female spores (Siddiqi *et al.*, 2000). The *SWITCH1* (*SWI1*) gene has been proposed to regulate the switch from mitosis to

meiosis because archesporial cells undergo an extra mitotic division before entering meiosis in ovules of swi1 mutant (Motamayor *et al.*, 2000). This gene was first thought to be involved exclusively in female meiosis (Motamayor *et al.*, 2000), but it was later demonstrated to be involved in sister chromatid cohesion during both male and female meiosis with clear differences (Mercier *et al.*, 2001). The *SYNAPSIS 1* (*SYN1*) gene (Bai *et al.*, 1999), or in another case called *DETERMINATE, INFERTILE 1* (*DIF1*) gene (Bhatt *et al.*, 1999), is a homolog of the *Schizosaccharomyces pombe REC8/RAD21* cohesin genes and is essential for meiotic chromosome segregation. Mutations in the *SYN1/DIF1* gene result in complete male and female sterility due to the disruption of chromosomal segregation, which is represented by univalent chromosomes and chromosome fragmentation at metaphase I, and acentric fragments and chromatin bridges in meiosis I and II. *AtDMC1*, the *Arabidopsis* homologue of the yeast *DMC1* gene, has been cloned and shown to be expressed during meiosis in both the male and female lineage (Klimyuk and Jones, 1997). Interestingly, *AtDMC1* is also expressed in the mitotically active cells from a suspension culture, indicating that it may also have a role in mitotic transcriptional regulation (Doutriaux *et al.*, 1998). In an *atdmc1* mutant generated by T-DNA insertion in the *AtDMC1* gene, in which its homozygotes fail to express *AtDMC1*, bivalent stabilization and chromosome segregation during both male and female meiosis are severely disturbed, resulting in dramatically reduced fertility (Couteau *et al.*, 1999). The correlation of the sterility phenotype with the abnormal chromosome behaviour in meiosis demonstrated that *AtDMC1* function is crucial for meiosis in *Arabidopsis*.

Arabidopsis genes specific for female meiosis have also been identified through the analysis of female sterile mutants such as *dyad* (Siddiqi *et al.*, 2000) and *antikevorkian* (*akv*) (Yang and Sundaresan, 2000). In *dyad* mutant, plants have normal pollen but ovules are arrested at the end of meiosis I after a normal reductional division. The *DYAD* gene may therefore encode a positive regulator required for entry into meiosis II (Siddiqi *et al.*, 2000). Unlike the meiosis-associated genes mentioned above,

AKV is a gene involved in post-meiotic events. The normal products of meiosis are four megaspores of which only the chalazal most spore becomes functional and the other three degenerate. In plants carrying the *akv* mutation, all four megaspores can develop equally and form multiple embryo sac (Yang and Sundaresan, 2000). AKV is thus revealed to be a locus essential for triggering the degeneration of the three unused spores to keep single embryo sac per ovule.

1.3.3.5 Megagametogenesis

The first mitosis of the functional megaspore marks the beginning of megagametogenesis. Currently, there appears to be that no gene specific for megagametogenesis has been isolated In *Arabidopsis*. The only cloned gene necessary for megagametophyte development is *PROLIFERA* (*PRL*) (Springer *et al.*, 1995). With a role in DNA replication, *PRL* is required early in megagametophyte development (before second mitotic division) but it also has a more general role in dividing cells throughout the plant. However, some mutations affecting nuclear division, polar nuclear fusion, cell specification and cellularization associated with megagametogenesis have been reported (Drews *et al.*, 1998; Schneitz, 1999; Yang and Sundaresan, 2000). Mutants such as *female gametophyte 2* (*fem2*), *fem3* (Drews *et al.*, 1998), *gametophytic factor* (*gf*) (Christensen *et al.*, 1997), *gametophytic factor 4* (*gfa4*), *gfa5* (Feldmann *et al.*, 1997) *andarta* (*ada*) and *tistrya* (*tya*) (Howden *et al.*, 1998) have defects in the first division. These mutants are affected at the one-nucleate stage, suggesting that haploid-expressed genes are required very early in megagametogenesis. Mutant *cell division cycle 16* (*cdc16*) (Yang and Sundaresan, 2000) has defects in second division while mutant *hadad* (*hdd*) (Moore *et al.*, 1997) showed defects in all three divisions. Genes involved in these processes could be part of the pathways including nuclear division and polar nuclei migration. In addition, some other mutants like *gfa2*, *gfa3* and *gfa7* have defects in post-mitosis development. In

these mutants, polar nuclei migration is normal but polar nuclei fail to fuse (Drews *et al.*, 1998). The isolation of genes mediating these steps would be very important for the elucidation of the mechanism underlying embryo sac formation.

1.3.3.6 Interactions between the sporophytic tissues of ovule and the megasporophyte or the megagametophyte

The questions about the influence of sporophytic tissues on megasporophyte and megagametophyte development have been asked but no direct answers have been obtained due to the unclear nature of this type of interaction. However, the studies of several mutants affecting both the sporophytic parts of ovule and the embryo sac have given some hints that the sporophytic tissues such as integuments play a role in the formation of the megasporocyte and/or megagametophyte (Gasser *et al.*, 1998; Yang and Sundaresan, 2000). Mutants of *ant*, *hll*, *bel1* and *sin1* all have either missing or abnormal integuments and early arrest of embryo sac development. Interestingly, in every case in which integuments do not enclose the nucellus, an embryo fails to form. However, some mutants have aberrant embryo sacs despite the presence of normal integuments. One such example is the *sterile apetala* (*sap*) mutant, which shows severe aberrations in inflorescence, flower and ovule development (Byzova *et al.*, 1999). In this mutant, the integument development is normal but megasporogenesis is arrested during or just after the first meiotic division. These results have led to the hypothesis that failure in embryo sac formation may be an indirect result of the absence of integuments and the integuments simply play a supportive role in promoting meiosis.

1.4 Approaches for the study of female gametophyte development

1.4.1 Morphological observation of wild-type ovule and female gametophyte development

Over the last decade, detailed morphological observations have provided vast amount of information on the whole process of wild-type ovule development in *Arabidopsis*. This has led to the establishment of a more accurate definition of developmental stages in ovule ontogenesis (Mansfield and Bowman, 1994; Schneitz *et al.*, 1995), and furthermore, in megagametogenesis (Christensen *et al.*, 1997). All these works are necessary prerequisites for the further systematic genetic and molecular analysis of this process because the staging system provides a framework for mutant analysis and other marker or reporter based analysis (e.g., transposon or T-DNA tagging) for gene cloning.

Morphogenesis of wild-type *Arabidopsis* ovules was first examined from initiation to maturation by scanning electron microscopy (Robinson-Beers *et al.*, 1992). The complete process was later examined in more details by light microscopic observation of Mayer's hemalum stained and optically cleared whole-mount ovules, supplemented by confocal laser scanning microscopy of propidium iodide-stained whole-mount ovules (Schneitz *et al.*, 1995).

Arabidopsis megagametogenesis has commonly been examined by light microscopy of paraffin-sectioned or resin-sectioned materials stained with toluidine blue or periodic acid-Schiff's staining (Misra, 1962; Webb and Gunning, 1990; Mansfield *et al.*, 1991), or using cleared ovules under differential interference contrast (DIC, or Nomaski) optics (Webb and Gunning, 1994; Schneitz *et al.*, 1995). Recently, the development of a rapid procedure using confocal laser scanning microscopy (CLSM) has further facilitated the analysis of female gametophyte development in

Arabidopsis (Christensen *et al.*, 1997). CLSM is particularly suitable for the observation of large number of samples because specimen preparation does not require time-consuming physical sectioning and is also faster than the tissue staining and clearing procedures used with DIC microscopy. This advantage has allowed the determination of eight morphologically distinct stages in megagametogenesis: first mitosis, central vacuole formation, second mitosis, third mitosis, cellularization, polar nuclei fusion, antipodal degeneration and synergid degeneration (Christensen *et al.*, 1997).

1.4.2 Female gametophyte mutations

Genetic approaches based on mutation of genes involved in pathway that results in certain loss-of-function phenotype have been a powerful tool for identifying important genes involved in flower, floral meristem, ovule, and embryo development (Ma, 1997; Sessions *et al.*, 1998)(Gasser *et al.*, 1998; Mayer and Jurgens, 1998). Progress has also been made in identifying mutations affecting specific steps of female gametophyte development and function (Drews *et al.*, 1998). In *Arabidopsis*, female gametophyte mutants have been generated by a variety of means, from the relatively 'ancient' methods such as X-ray irradiation and ethyl methanesulfonate (EMS) induction to the more recent methods based on transposon or T-DNA insertion. For example, the *Gf* mutant (Redei, 1965; Christensen *et al.*, 1997) was obtained from X-ray irradiation. The mutants *fis1*, *fis2*, *fis3* (Chaudhury *et al.*, 1997), *fem1* (Christensen *et al.*, 1998) and *fie* (Ohad *et al.*, 1996) were isolated from the screen of EMS-mutagenized lines. The mutants *fem2*, *fem3*, *fem4* (Christensen *et al.*, 1998) and *gfa1- gfa7* (Feldmann *et al.*, 1997) were isolated from the screen of T-DNA-mutagenized lines, while some mutants, such as *hadad* (Moore *et al.*, 1997) and *mea* (Grossniklaus *et al.*, 1998b), were generated by Ds insertion.

Mutations affecting the female gemetophyte can be identified as lethals in which female gametophytes harboring the mutation are defective and therefore they either

abort development or are nonfunctional. In theory, female gametophyte mutants can always be identified by reduced seed set. A mutation affecting female gametophyte development or function fails to transmit through the female gametophyte and consequently the ovule arrests. In a heterozygous plant, one half of the ovules will receive the mutant allele that results in a semisterile phenotype (siliques containing 50% normal seeds and 50% desiccated ovules). However, reduced seed set can also be caused by some other factors, including adverse environmental conditions (e.g., high temperature and water stress), chromosomal rearrangement (e.g., reciprocal translocations and large inversions) and some sporophytic mutations (e.g., female sterile mutation with partial penetrance) (Christensen *et al.*, 1998).

To effectively identify female gametophyte mutations, a second type of screen, segregation distortion, can be combined with a seed set screen. In segregation distortion screens, either gametophyte mutation linked visible markers or gametophyte mutation linked molecular markers have been used for the detection of altered segregation (Drews *et al.*, 1998). For example, in dealing with T-DNA-mutagenized lines harboring T-DNA with a *nptII* gene conferring kanamycin resistance, Feldmann et al. (1997) demonstrated that the exceptional ratio of kanamycin-resistant (KanR) to kanamycin-sensitive (KanS) seedlings can be used to identify gametophyte mutations. If a plant carrying a T-DNA insert disrupting a gene required for female gametophytic growth is heterozygous for the mutation, its progeny will show a KanR:KanS ratio of 1:1 (if the mutation is fully penetrant and does not affect the male gemetophyte). By using these criteria, many female gametophyte mutants have been isolated (Drews *et al.*, 1998). All these mutants are valuable materials for the isolation of genes responsible for female gametophyte development.

1.4.3 Enhancer and gene traps: new systems for identification of tissue- or development-specific genes and enhancers

Enhancer traps and gene traps were more recently developed systems for detecting genes and cis-regulatory elements. They are reporter gene constructs that can respond to cis-acting transcriptional signals when inserted in the genome of certain organism. Construction of the trap population involves generating a large number of individuals that have the reporter gene integrated at different sites throughout the genome. The trap population can be used for two purposes: first, for examining the pattern of reporter gene expression, and second, for screening the mutant phenotypes caused by insertion. Since the random insertion of the enhancer trap or gene trap element in the genome could lead to certain mutation phenotype if the insert disrupts a key gene, just like normal transposon or T-DNA based insertional mutagenesis (Bancroft *et al.*, 1993; Azpiroz-Leehan and Feldmann, 1997), the trap population can be directly used for screening phenotypically recognizable mutants.

However, additionally, a more useful aspect of the trap population is that it permits the identification of genes, which could have been missed in conventional mutagenesis screens, by their patterns of expression. The conventional mutagenesis approaches are dependent on the presence of a recognizable phenotype when a particular gene is mutated. These approaches have at least two limitations. First, many genes play multiple roles and phenotypic screens may fail to detect such genes that exhibit discrete functions at different stages in development. Second, if a gene is functionally redundant, disruption of the gene may result in only subtle phenotype or no phenotype at all unless all genes specify that function are mutated simultaneously. In *Arabidopsis*, there are many examples of functional redundancy among closely related genes (Kempin *et al.*, 1995; Pelaz *et al.*, 2000; Ferrandiz *et al.*, 2000; Liljegren *et al.*, 2000). For example, in the SEPALLATA family of MADS box transcription factors, single gene mutation of *sep1*, *sep2* and *sep3* produced only subtle phenotypes.

However, *sep1 sep2 sep3* triple mutants revealed striking phenotypes (Pelaz *et al.*, 2000). These functionally redundant genes are found on the segmental duplications which comprise 58% of the *Arabidopsis* genome (The *Arabidopsis* Genome Initiative, 2000). Such genes can not be identified in forward genetic screens. These limitations could be overcome by enhancer trap or gene trap screens because lines are selected based on expression pattern of reporter gene not mutant phenotype. Even if the inserted trap disrupt an essential gene that has a lethal effect, the plant will normally still be viable when the insert is hemizygous (heterozygous) and the reporter gene expression pattern can be detected due to the dominant nature of the trap element.

The concept of enhancer trap and gene trap have been successfully used in *Drosophila* and in mouse, leading to the isolation of genes important in development whose phenotypes would have been missed in conventional mutagenesis screens. Enhancer or gene trap has been constructed based on the P-element in *Drosophila* (O'Kane and Gehring, 1987; Wilson *et al.*, 1989; Bellen *et al.*, 1989) and the retrovirus in mouse (Gossler *et al.*, 1989; Skarnes, 1990). A bacterial β-galactosidase (*lacZ*) gene was used as a reporter gene in the trap constructs in both organisms. In the enhancer trap for *Drosophila*, the *lacZ* gene is placed under the control of a weak, constitutive P-transposase promoter. This fusion gene is part of the P-element that can be integrated at different locations in the *Drosophila* genome. The weak P-transposase promoter could be influenced by nearby genomic regulatory elements that impose temporal and spatial regulations on the fusion gene and such effect can be visualized by *lacZ* expression patterns (Bellen *et al.*, 1989). Similarly, in the enhancer trap for mouse, a minimal promoter from the mouse heat-shock protein 68 (*hsp68*) gene is used to fuse with the *lacZ* gene to form a trap unit (Gossler *et al.*, 1989). The gene trap for mouse contains a promoter less *lacZ* gene inserted in frame into the exon of the *En-2* gene such that a splice accepter is placed at the 5' end of *lacZ* (Gossler *et al.*, 1989). Integration of such construct into introns of genes in the right orientation could create a

functional LacZ fusion protein that can be detected by X-Gal staining to localize the tissues where the tagged gene expresses. Some new developments in using enhancer or gene trap for enhancer and gene hunting in *Drosophila* and mouse have also been reported (Mollereau *et al.*, 2000; Gates and Thummel, 2000; Lukacsovich *et al.*, 2001; Stanford *et al.*, 2001).

The same strategies of using enhancer trap and/or gene trap approaches in detecting cis-elements and genes have been employed in plants (Springer, 2000). Progress has been made in *Arabidopsis* (see chapter 3) and rice (Chin *et al.*, 1999). The establishment of such a gene detector system in plant relies on the progress of the following aspects: gene transfer techniques, transposon/T-DNA work and reporter genes, because a gene detector construct needs to be efficiently delivered into plant and be randomly inserted into the host genome before any target gene action can be revealed by a reporter gene. For transferring a gene into plants, the well established *Agrobacterium*-mediated transformation procedures have been proven to be very efficient to generate large transgenic populations in both dicot plants like *Arabidopsis* (see chapter 5) and monocot plants such as rice (Hiei et al., 1997). To insert a gene detector construct into a plant genome, both maize transposon systems (Osborne and Baker, 1995) and T-DNA vector (Koncz *et al.*, 1989; Feldmann, 1991) have been used as carriers. They have been extensively used in insertional mutagenesis for functional genomics study.

As for reporter genes in plant, the β-glucuronidase (GUS) gene (Jefferson, 1989) is the most widely used one due to its reliability and high sensitivity. In some cases, the gene coding for neomycin phosphotransferase II (*nptII*) has also been used as a reporter gene (Koncz *et al.*, 1989; Babiychuk *et al.*, 1997). The green fluorescence protein (GFP) gene (Sheen *et al.*, 1995), with its advantage of detecting gene expression without destroying the plant, has been tested as a reporter gene in an

enhancer trap system in rice and shown to be a good alternative to the GUS gene (X. Fu and A. Kilian, unpublished data).

In plants, the gene traps are designed to create fusion transcripts with the target gene. The GUS gene is able to support gene fusions and have been used widely to form gene fusion trap in two types: promoter trap and exon trap. The promoter trap consists of a promoterless *gusA* gene at the T-DNA right border or left border and is expressed when inserted downstream of a chromosomal gene promoter(Kertbundit *et al.*, 1991; Lindsey *et al.*, 1993; Topping and Lindsey, 1995). The exon trap, which is very similar to the one used in mouse, has consensus splicing accepter sites upstream of the *gusA* gene. When the exon trap inserted in an intron within a transcribed region, transcriptional fusion could be created to report the existence of a gene (Sundaresan *et al.*, 1995). Since it enables reporter gene to create transcriptional fusions at various locations within a gene, exon trap is therefore more versatile.

The enhancer trap systems used in plants are developed to detect cis-regulatory elements in the genome, which are normally able to activate transcription of certain gene from a distance in an orientation independent manner. A well-known enhancer trap system is the one established by Sundaresan *et al.* (1995) based on the two component Ac/Ds transposable elements (Bancroft *et al.*, 1992). The system uses a modified Ds element carrying a β-glucuronidase gene fused to a 35S minimal promoter as enhancer trap (Figure 1.4). Various starter lines, each containing a single stable DsE insertion, are generated by *Agrobacterium*-mediated transformation. Transposition of the DsE is achieved by crossing the DsE starter lines with lines expressing the Ac transposase. Since parental Ac transposase gene is linked to a gene coding for the indole acetic acid hydrolase (IAAH), which confers sensitivity to naphthalene acetamide (NAM), progenies that are free of transposase activity can be selected so that each selected DsE insertion is stable. The *IAAH* gene is also linked to the DsE in the T-DNA donor site to act as a counter-selectable marker for eliminating

the progenies retaining the T-DNA donor site in the presence of NAM. At the same time, plants with a transposition event are selected by the kanamycin resistance gene *nptII* contained in the same DsE (Figure 1.4). This selection procedure ensures that only the plants with transposition to more distal site can survive. In this way, the genome can be saturated with random DsE insertions.

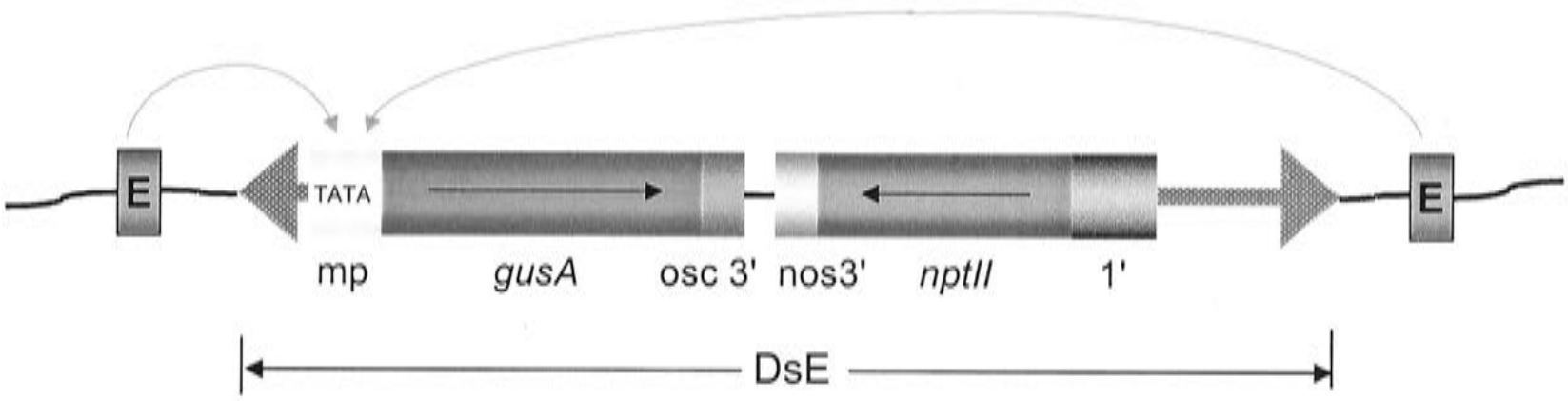


Figure 1.4 Expression of the GUS reporter gene from the enhancer trap Ds element (DsE) inserted near a chromosomal enhancer (modified from Sundaresan *et al.*, 1995).

Putative enhancers (E) can activate GUS expression from both the upstream and the downstream regions of the DsE insertion site in the genome. Parts in the DsE are: *gusA*, β-glucuronidase gene; mp, 35S minimal promoter; *nptII*, neomycin phosphotransferase II gene; nos 3', terminator of nopaline synthase gene; ocs 3', terminator of octopine synthase gene; 1', 1' promoter. The big broken left-right arrow represents the Ds element.

As described in 1.2.3, very few genes expressed in the female gametophyte have been identified and isolated. Identifying genes expressed in specific cells of the embryo sac has been hampered by its relatively inaccessible nature and lethality involved in the female mutants. However, it is more feasible than ever before to dissect female gametophyte development through examining large collection of gene trap and enhancer trap lines in *Arabidopsis*. It has already been shown that some of such lines exhibit reporter gene expression in specific cells in the embryo sac (Grossniklaus *et al.*, 1995). With the continuation of this screening work, more lines representing reporter

gene expression in different stages of the megagemetogenesis should be identified. These lines are very useful as markers for the different megagametophyte cell types. The use of enhancer trap lines that label one cell or a small group of cells allows these cells to be followed at different times in development and in different mutant background. And moreover, the trapped genes or enhancers could be isolated and they could serve as entry points into the gene regulatory circuitry in female gametophyte development (Drews *et al.*, 1998).

1.5 Aims of the project

With an ultimate goal on the introduction of an apomictic mechanism into economically important crops such as rice, we need to have a clear understanding of the mechanisms behind the plant reproductive systems as a first step. This involves the identification of key genes and regulatory elements associated with the plant reproductive process, the revealing of the network of these gene interactions and the manipulation of such genes. Although this review is selective and does not attempt a complete overview of the many publications and presentations in the area, I hope it will serve to introduce the readers of this thesis into the following chapters. By using the model plant *Arabidopsis thaliana*, I aimed mainly on the following aspects for my PhD project:

1. Identification of female gametophyte-specific regulatory elements (enhancers) and genes from *Arabidopsis* by using enhancer trap lines based on transposon tagging.

2. Development of a reliable enhancer testing system in *Arabidopsis* for testing DNA sequences with certain enhancer activity, especially those with tissue-specific activities in flower.

3. Exploration of the possibilities of using the cloned enhancer(s) for controlling specific gene expression in female gametophyte as a mean of manipulating gene activity in female reproductive cells.

CHAPTER 2

MATERIALS AND METHODS

2.1 Nucleic acid extraction

2.1.1 Plasmid DNA preparation

Plasmid DNA extraction was routinely done as described in Sambrook *et al.* (Sambrook *et al.*, 1989). For small-scale preparations, a single bacterial colony was inoculated into 3 ml of LB medium containing the appropriate antibiotics in a loosely capped 15-ml tube. The culture was incubated at 37°C with vigorous shaking (about 200 rpm) for overnight (16-18 hours). 1-1.5 ml of the culture was then transferred to a 1.5 ml microfuge tube and centrifuged at 12,000*g* for 30 sec. After discarding the medium, the bacterial pellet was resuspended in 100 µl of freshly prepared Solution I (50mM glucose, 25mMTris·Cl, 10mM EDTA, pH8.0) and 200 µl of Solution II (0.2 M NaOH, 1% SDS) was added. The contents was mixed by inverting the tube rapidly several times and stored on ice for 5 min. 150µl of Solution III (prepared by mixing 60 ml of 5M potassium acetate, 11.5ml of glacial acetic acid and 28.5ml of H_2O) was then added and gently mixed. After keeping on ice for 5 min, the tube was centrifuged at 12,000*g* for 5 min. The supernatant was transferred to a new tube containing 400µl of phenol:chloroform (1:1) and mixed by vortexing for 30 sec. After centrifuging at 12,000*g* for 2 min, the supernatant was transferred to a new tube containing 900µl of 100% ethanol to precipitate plasmid DNA. The DNA pellet was washed once with 1ml of 70% ethanol and briefly dried in the air and dissolved in 20-30µl of TE (pH 8.0) containing 20µg/ml RNAase. This plasmid DNA solution was stored at either -20°C or 4°C.

2.1.2 Plant genomic DNA preparation

Genomic DNA from *Arabidopsis* was prepared using the CTAB method (Murray and Thompson, 1980) with minor changes. Approximately 4g of plant tissues (leaf, stem, flower etc) were collected and ground to fine powder in liquid nitrogen in a mortar. The powder was transferred into a 50 ml centrifuge tube containing 14 ml of Buffer S (110mM Tris·Cl pH8.0, 55mM EDTA, 1.54M NaCl, 1.1% CTAB, pre-warmed to 65°C) and mixed well. 1.4ml of 20% SDS solution was then added and mixed gently. The mixture was incubated at 65°C for 30-60 min with gentle shaking for cell lysis. 1ml of chloroform was added when the tube was cooled down to RT and the sample was shaken vigorously for 30 sec. After centrifuging at 12,000*g* for 10 min, the supernatant was transferred to a new tube and 0.6 volume of isopropanol was added and mixed thoroughly. The precipitated DNA was collected by spinning at 12,000*g* for 5 min and dissolved in 400µl of TE (pH8.0) containing 100µg/ml DNAase-free RNAase. The DNA solution was incubated at 55°C to remove RNA and then extracted once with phenol:chloroform (1:1). DNA was precipitated again by adding 40µl of 3M sodium acetate (pH 5.2) and 1 ml of ethanol. The precipitated DNA was hooked out by a glass hook, rinsed with 70% ethanol, dried and dissolved in 200µl of TE (pH 8.0).

For mini-preparation of genomic DNA for PCR analysis, a simple method as described by Zheng *et al.* (Zheng *et al.*, 1995) was used. A healthy *Arabidopsis* leaf (about 1.5mg) was placed in a 1.5ml microfuge tube containing 50µl of DNA extraction buffer (50mM Tris·Cl pH8.0, 25mM EDTA, 300mM NaCl, 10% SDS) and homogenized on ice with a thick polished glass rod. The liquid was extracted once with 50 µl of chloroform and the aqueous phase was transferred to a fresh tube containing 100 µl of ethanol for DNA precipitation. The precipitated DNA was collected by spinning at 13,000 rpm for 5 min in a microfuge. The DNA pellet was washed once with 70% ethanol, briefly dried and dissolved in 20µl of TE.

2.1.3 Total RNA and mRNA preparation

Total RNA was extracted from different tissues of *Arabidopsis*, such as leaf, stem and flower, using TRIzol reagent (Life Technologies). 50mg of tissue was placed in a microfuge tube and homogenized with a polished glass rod after adding 0.5ml of TRIzol Reagent (a mono-phase solution containing phenol and guanidine isothiocyanate). 100µl of chloroform was then added and mixed by vigorous shaking for 30 sec. The sample was centrifuged at 12,000*g* for 10 min and the upper colorless aqueous phase was transferred to a fresh tube. RNA was precipitated by adding 0.3ml of isopropanol, mixing well and keeping at RT for 10 min. The RNA precipitate was pelleted by spinning at 12,000*g* for 10 min, washed twice with 75% ethanol, briefly dried and dissolved in 20µl of RNAse free DEPC-treated water. The RNA solution was incubated at 60°C for 10 min before being stored at -80°C.

mRNA was prepared from total RNA using Dynabeads Oligo $(dT)_{25}$. (Dynal A. S., Norway) with the Dynabeads mRNA Purification kit. Total RNA prepared by TRIzol Reagent was used for mRNA isolation. For conditioning of Dynabeads Oligo $(dT)_{25}$, 200µl (1mg) of Dynabeads Oligo $(dT)_{25}$ was transferred from the stock tube suspension to a new 0.5ml tube placed in a Dynal MPC-E magnetic particle concentrator to set the Dynabeads. After the supernatant was removed, the tube was removed from the Dynal MPC. The Dynabeads was then washed twice in 100µl of 2× Binding buffer (20mM Tris·Cl pH7.5, 1.0M LiCl, 2mM EDTA) using the Dynal MPC and resuspended in 100µl of 2× Binding buffer. For isolation of mRNA, total RNA (up to 100µg) was diluted in 100µl of DEPC-treated water, heated to 65°C for 2 min, and mixed with the100µl conditioned Dynabeads. After keeping at RT for 5 min to allow the annealing, the tube was placed in Dynal MPC for 30 sec and the supernatant was removed and saved in a new tube for next extraction. The tube was then transferred to a normal rack and the

Dynabeads with mRNA attached was washed twice with 200µl of washing buffer (10mM Tris·Cl pH8.0, 0.15M LiCl, 1mM EDTA) using Dynal MPC. For elution of mRNA from the Dynabeads, 20µl of Elution solution (2mM EDTA pH 8.0) was added, mixed and heated to 65°C for 2 min. The tube was then placed in the Dynal MPC and the supernatant containing mRNA was transferred to a new tube on ice. After washing the Dynabeads once in the 2× binding buffer, this extraction process were repeated for three times with the saved total RNA solution and the elution solution containing mRNA to maximize the mRNA extraction. The mRNA solutions were store at -80°C for further use.

2.1.4 Quantitation of DNA and RNA

The amount of DNA or RNA in a preparation was routinely determined by measuring OD_{260} on a spectrophotometer. An OD_{26} of 1.0 equals to approximately 50µg/ml for double-stranded DNA, 33µg/ml for single-stranded DNA and 40µg/ml for single-stranded RNA. The concentration of DNA or RNA solution was calculated accordingly.

2.2 DNA manipulation and cloning

The techniques used for manipulating DNA were generally according to Sambrook *et al.* (Sambrook *et al.*, 1989). Restriction endonucleases were mainly purchased from New England Biolabs and the digestions were normally carried out in 20µl volume using the buffers provided with the enzymes. Blunt-ended DNA fragments were generated either by filling in the 3' recessed termini using Klenow fragment of *E. coli* DNA polymerase New England Biolabs), or by polishing the protruding 5' and 3' termini using T4 DNA polymerase (New England Biolabs). Dephosphorylation of 5' phosphate from DNA was performed using shrimp alkaline phosphatase (Boehringer Mannheim). Ligation of DNA fragments was carried out using T4 DNA ligase (New England Biolabs).

2.3 Polymerase chain reaction (PCR)

2.3.1 Standard PCR

Standard PCR for detection of genes in plant genome or confirmation of plasmid existence in *Agrobacterium* strains was routinely performed based on the method described by Saiki *et al.* (Saiki *et al.*, 1988). The PCR reaction generally contained DNA template (either purified DNA or cell lysis), 1×PCR buffer (10mM Tris·Cl pH 8.3, 50mM KCl, 0.01% gelatin), 1.5mM $MgCl_2$, 800µM dNTPs, 0.5µM of each of the forward primer and reverse primer, and 0.2 units of Taq polymerase. PCR program was normally set at 30 cycles of 30 sec denaturation at 94°C, 30 sec annealing at 50-60°C and 1-2 min extension at 72°C, and run on FTS-960 thermal sequencer or PC-960 thermal cycler (Corbett Research).

2.3.2 Thermal Asymmetric Interlaced (TAIL-) PCR

TAIL-PCR was performed as described by Liu *et al.* (1995). A schematic chart depicting the main steps involved in Inverse PCR is shown in Fig. 2.1. Nested primers Ds5-1, Ds5-2, Ds5-3 and Ds3-1, Ds3-2, Ds3-3, which are specific to the 5' and 3' ends respectively, were designed by Dr Ueli Grossniklaus. The arbitrary primer AD2 was the same as described by *Liu et al.* (1995). Primary TAIL-PCR mixture (20µl) contained 20-50ng of genomic DNA from enhancer trap line, 1×PCR buffer (as in 2.3.1), 200µM of each of the dNTPs, 0.2µM of Ds5-1 or Ds3-1, 0.2µM of AD2, and 1 unit of Taq polymerase. The primary TAIL-PCR reaction was executed on GeneAmp PCR system 9600 with 5 high stringency cycles (30 sec at 94°C, 1 min at 62°C, 2.5 min at 72°C), followed by 1 low stringency cycle (30 sec at 94°C, 3 min at 62°C, 3 min ramping to 72°C, 2.5 min at 72°C), then followed by 15 super cycles (10 sec at 94°C, 1 min at 68°C, 2.5 min at 72°C, 10 sec at 94°C, 1 min at 68°C, 2.5 min at 72°C, 10 sec at 94°C, 1 min at 44°C, 2.5 min at 72°C). Aliquot (1µl) of 50-fold dilutions (in H_2O) of the primary

PCR products was added to secondary TAIL-PCR mixture (20 µl) containing 1×PCR buffer, 200µM of each of the dNTPs, 0.2µM of Ds5-2 or Ds3-2, 0.2µM of AD2, and 0.5 unit of Taq polymerase.

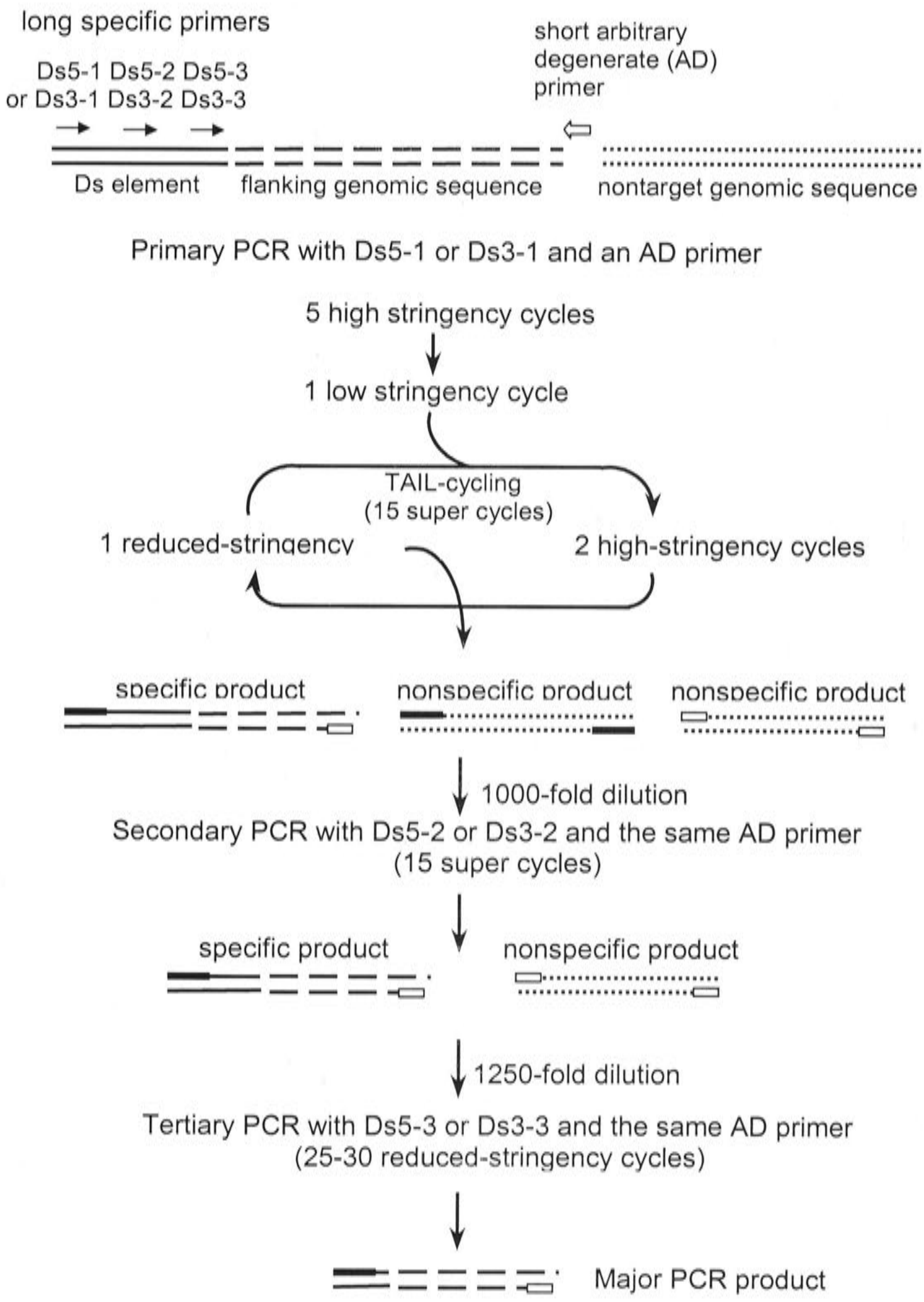


Figure 2.1 TAIL-PCR procedure for specific amplification of 5' and 3' genomic sequences flanking a Ds element.

Three PCR reactions are executed sequentially to amplify target sequences using nested primers specific to either 5' or 3' end of Ds element together with an short arbitrary primer (AD). The nested specific primers and the AD primer are depicted as filled arrows and open arrow, respectively. One or more sites within the Ds flanking sequences are adapted for annealing to the AD primer through a special low-stringency cycle. After creation of annealing sites for AD primer, high temperature annealing is applied to favor the specific primer. This asymmetric (uneven) treatment results in a linear amplification of target fragments. (Modified from Liu et al., 1995).

The secondary TAIL-PCR reaction was executed with 15 super cycles of 10 sec at 94°C, 1 min at 64°C, 2.5 min at 72°C, 10 sec at 94°C, 1 min at 64°C, 2.5 min at 72°C, 10 sec at 94°C, 1 min at 44°C, 2.5 min at 72°C. 2μl aliquot of 50-fold dilutions of the secondary PCR products was then applied to tertiary TAIL-PCR mixture (50μl) with the same components and concentrations except for the replacement of Ds5-2 or Ds3-2 by Ds5-3 or Ds3-3. The tertiary reaction was run for 25-30 reduced-stringency cycles of 15 sec at 94°C, 1 min at 44°C, 2.5 min at 72°C to re-amplify the Ds-specific products. Amplified tertiary PCR products were analyzed by agarose gel electrophoresis. Ds-specific fragments were recovered from the gel by either Qiagen Gel Extraction kit or GlasPac/GS kit (National Scientific Supply Company) according to manufacturer's instructions and sequenced directly as in 2.4.

2.3.3 Inverse PCR

Inverse PCR was carried out basically according to Ochman *et al.* (Ochman *et al.*, 1988; Ochman *et al.*, 1990; Ochman *et al.*, 1993). A schematic chart depicting the main steps involved in Inverse PCR is shown in Fig. 2.2. In general, It has three main steps: digestion of genomic DNA with appropriate restriction enzyme, circularization of the digested DNA fragments and PCR amplification. The choice of restriction enzyme can be established by examining the sequence of the known region (or called core region)

or by determining the cleavage sites and fragment lengths by Southern analysis. The procedure is as follows: 0.5-1μg of genomic DNA was digested with appropriate restriction enzyme in 30μl volume for overnight. The overnight digest was extracted twice with phenol:chloroform (1:1) (or alternatively purified by Qiagen spin column) to remove the restriction enzyme.

For circularization, DNA fragments were precipitated by ethanol and resuspended in 250-500μl of 1× DNA ligase buffer (50mM Tris·Cl pH 7.5, 10mM $MgCl_2$, 10mM dithiothreitol, 1mM ATP, 25μg/ml bovine serum) depending on the initial amount of DNA used (usually 2 μg/ml DNA is the suitable concentration for the formation of monomeric DNA circles). 2-5μl of T4 DNA ligase (400 units/μl, New England Biolabs) was added and the Ligation was carried out at 16°C for overnight. The ligated sample was heated to 75°C for 10 min to inactivate the ligase and the DNA was precipitated by ethanol again. After being washed once with 70% ethanol and briefly dried, the DNA pellet (usually invisible) was dissolved in 20μl of TE (pH 8.0).

The PCR amplification was then performed by using the GeneAmp XL PCR kit (PE Applied Biosystems). Unlike the primers employed in standard PCR, primers for inverse PCR were designed such that they are complementary to the opposite strand and therefore the extension proceeds outward from the core region. Theoretically, any standard PCR procedure can be used for the amplification of inverse PCR fragments. The amplification with GeneAmp XL Kit, however, is very good especially for obtaining long PCR products (ranging from 5 to 40kb) with high fidelity (Cheng *et al.*, 1994; Stewart *et al.*, 1995). A typical PCR mixture and conditions for GeneAmp XL system was as follows: 6μl of above circularized DNA, 15μl of 3.3× XL Buffer II, 4μl of 10mM dNTPs, 7μl of 25mM $Mg(OAc)_2$, 1μl of 10μM primer1, 1μl of 10μM primer2 and 16μl of H_2O (50μl in total) was mixed in a thin wall PCR tube and heated to 94°C on GeneAmp PCR system 9600 (PE Applied Biosystems) before adding 1-2μl (2-4 units) of *rTth* DNA Polymerase XL. PCR was then run for 30-35 cycles of denaturation at

94°C for 30 sec, primer annealing at 58-60°C for 1min, and extension at 72°C for 5-10 min depending on the size of PCR product to be amplified. Specific PCR products were analyzed and sequenced as in 2.3.2.

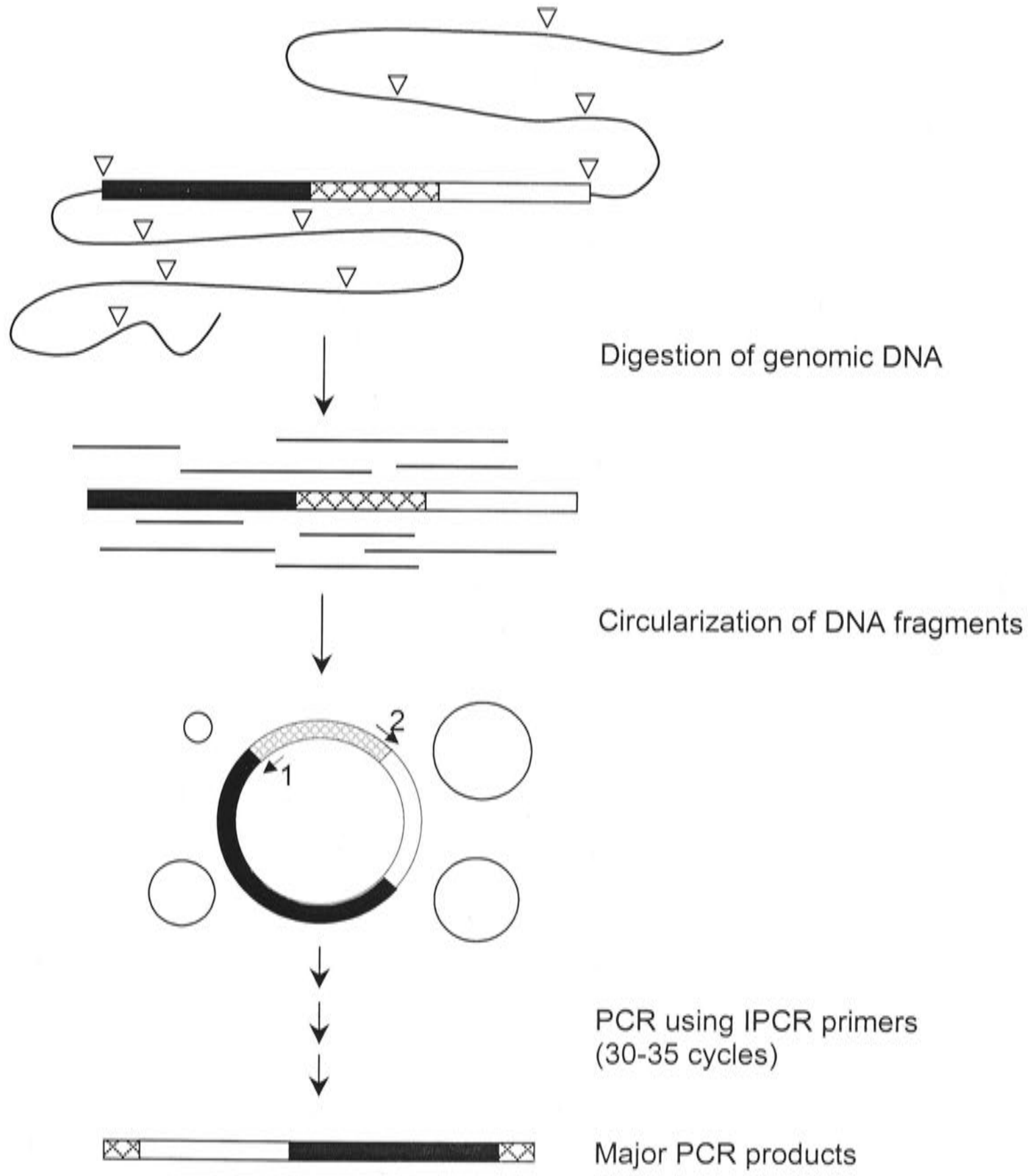


Figure 2.2 Schematic representation of the inverse PCR procedure.

The core region is depicted as an outlined diamond box. The black and open boxes represent the upstream and downstream flanking regions, respectively. The recognition sites of restriction enzyme are denoted by open triangles. Oligonucleotide primer1 and primer2 are designed based on the known

sequence in the core region and their directions are shown by arrows. (Modified from Ochman *et al.*, 1990)

2.4 Oligo synthesis and DNA sequencing

All oligonucleotides for PCR and sequencing were synthesized either in Division of Plant Industry, CSIRO or GENSET Pacific Company.

DNA sequencing was based on the dideoxynucleotide chain termination method (Sanger *et al.*, 1977) but using automatic sequencing machine. Plasmid DNA and PCR fragments were used as sequencing template. Sequencing reactions were carried out using the Big Dye Terminator Ready Reaction Mix (PE Applied Biosystems) with cycle sequencing program set at 25-30 cycles of 10 sec at 95°C, 5 sec at 50°C, and 4 min at 60°C. The products were then precipitated by ethanol and sent to Australian Genome Research Facility, University of Queensland for gel separation. Sequencing results were obtained directly from AGRF web site as FTP files and viewed using Chromas software version 1.45 (Conor McCarthy, School of Health Science, Griffith University)

2.5 Nucleic acid hybridization

2.5.1 Southern hybridization

Southern hybridization of *Arabidopsis* genomic DNA was basically as described by Southern (Southern, 1975) but positively charged nylon membranes were used. About 5µg of *Arabidopsis* genomic DNA was digested completely with one or more restriction enzymes at 37°C (or other appropriate temperatures) for overnight and separated by electrophoresis on 1% (or other appropriate concentration depending on the size of the fragment to be detected) agarose gel. After electrophoresis, the gel was stained in 0.5µg/ml ethidium bromide for 30 min and photographed. For transferring the

fractionated DNA fragments from the gel to a positively charged nylon membrane (Boehringer Mannheim), the gel was treated in 0.25M HCl for 10 min (optional), rinsed with water, and denatured in 10 gel volume of denaturation/transfer solution (0.4M NaOH, 1.6M NaCl) for 30 min with gentle agitation. Capillary transfer was then carried out as described in Sambrook *et al.* (1989) using the same denaturation/transfer solution. After transferred for 16-24 hours, the membrane was peeled off from the gel, rinsed in 2× SSC, air dried and stored at 4°C for further use.

^{32}P-labeled DNA probe was prepared by first obtaining the DNA fragment to be used as probe through either restriction enzyme digestion or PCR amplification from a plasmid, and then labeling with α- ^{32}P-dATP (specific activity of 3000Ci/mmole) using GIGAPRIME DNA labeling kit (Bresatec). For hybridization, the nylon membrane containing the target DNA was wetted in 6× SSC and subjected to prehybridization in hybridization solution (1× HSB, 1× Denhardt's III reagent, 100μg/ml denatured Herring sperm DNA) for 4-16 hours at 65°C in a tube in XTRON HI 2002 hybridization oven (Bartelt Instruments). To start hybridization, the old hybridization solution was replaced by fresh hybridization solution (the volume was calculated by 0.2ml/cm^2 × total membrane area), and denatured probe (prepared by heating to 95°C for 2 min and Chilling on ice) was then added. Hybridization was carried out at 65°C for 16-20 hours. The hybridized membrane was washed in low-stringency solution (2×SSC, 0.1% SDS) for 15 min at RT, moderate-stringency solution (1×SSC, 0.5% SDS) for 30 min at RT, and high-stringency solution (0.1×SSC, 0.5% SDS) for 30 min at 65°C. Autoradiographic image was taken on Bio-Rad GS-250 Molecular Imager after 4-24 hours exposure using the Molecular Analyst software version 1.4 (with resolution set at 200μm).

2.5.2 Northern hybridization

Northern hybridization (Alwine *et al.*, 1977) of total RNA or mRNA from different tissues of *Arabidopsis* was mainly according to Sambrook *et al.* (1989). For separation of RNA, 5-15µg of total RNA or about 0.5µg of mRNA in 4.5µl volume was mixed with 2µl of 5× formaldehyde gel-running buffer, 3.5µl of 37% formaldehyde and 10µl of formamide, heated to 65°for 15 min, mixed with 2µl of gel loading buffer and loaded onto a denaturing agarose gel (1% agarose, 2.2M formaldehyde, 1× formaldehyde gel-running buffer). Electrophoresis was carried out in 1× formaldehyde gel-running buffer. After electrophoresis, the gel was rinsed in DEPC-water for 45 min to remove formaldehyde and then rinsed with 20×SSC for 5-10min. Transfer of RNA onto positively charged nylon membrane (Boehringer Mannheim) was done the same as Southern transfer except for using 20×SSC as transfer solution. After allowing about 20 hour's transfer, the membrane was removed from the gel, rinsed in 2×SSC for 5 min and air dried. For UV cross-linking, the membrane was placed on a transilluminator (254nm wavelength) for 3 min and then baked at 90°C for 15 min.

The hybridization procedure using ^{32}P-labeled DNA probe was the same as that used for Southern hybridization, except that the hybridization temperature was at 68°C. The membrane was then washed once in 2×SSC/0.1% SDS solution for 5 min at RT, twice in 0.2×SSC/0.1% SDS solution for 30 min at 68°C, and once in 0.1×SSC/0.1% SDS solution at 68°C (optional). Autoradiographic images were taken the same as in 2.5.1.

2.6 *In silico* work

2.6.1 Molecular map construction

All molecular maps were constructed using Vector NTI software (Infomax, USA).

2.6.2 BLAST sequence similarity searching

BLAST sequence similarity searching was used to find matching sequences and genes from databases (such as Genbank) after new sequences were obtained. BLAST stands for Basic Local Alignment Search Tool and was developed by et al. (Altschul *et al.*, 1990) and significantly improved later on (Altschul *et al.*, 1997). BLAST is best used for sequence similarity searching of unknown DNA or predicted protein (http://www.ncbi.nlm.nih.gov/BLAST or, for *Arabidopsis* sequences, a better site is: http://www.arabidopsis.org/Blast).

2.6.3 Sequence alignment

Sequence alignment was performed by using the Australian National Genomic Information Service (ANGIS) (http://www.angis.org.au). Two sequences were aligned using the **bestfit** program. For aligning more than two sequences, the program **pileup** was used.

2.7 Transformation

2.7.1 Transformation of bacteria

2.7.1.1 Transformation of Escherichia coli

All plasmids were maintained and propagated in *E. coli* strain DH5α (Hanahan, 1983). Plasmid was introduced into DH5α by electroporation (Dower *et al.*, 1988). To prepare electrocompetent DH5α cells, a single colony was inoculated into 3ml LB medium containing 10μg/ml of nalidixic acid and cultured at 37°C for overnight. 1ml of the overnight culture was then inoculated into 1L of 2YT medium and grown with vigorous shaking (240 rpm) until the OD_{600} reached 0.75-0.8. The cells were harvested by

centrifugation at 7000 rpm (Beckman JLA10.5 rotor) at 4°C for 10min, washed twice with up to 1L of ice-cold water and once with 100ml ice-cold 10% glycerol. The cells were then resuspended in 2 ml of ice-cold 10% glycerol, kept on ice for overnight (optional), aliquoted (150µl each), snap frozen in dry ice in ethanol and stored at -80°C.

For transformation, an aliquot of competent cells was thawed on ice and 30-50µl of cells was used for each plasmid sample. 10-50ng of plasmid or ligation mixture was used to mix with the competent cells on ice and then electroporated in a 0.2mm cuvette using the Bio-Rad Gene Pulser set at 2.45V, 200Ω and 25µF. After electroporation, 500µl of SOC medium was added and the cells were allowed to recover for 15-30 min at 37°C before being plated onto LB medium containing appropriate antibiotics. After being cultured at 37°C for 16-18 hours for ampicillin or kanamycin selection (for chloramphenicol selection at 10µg/ml, up to 24 hours culture is needed), an individual colony was picked and plasmid was extracted as described in 2.1.1 and identified by restriction enzyme digestion.

2.7.1.2 Transformation of Agrobacterium tumefaciens

The procedure for transformation of *Agrobacterium tumefaciens* is similar to that of *E coli* with some changes. For all plant transformation experiments, strain EHA105 (Hood *et al.*, 1993) was used. To prepare electrocompetent cells, a single EHA105 colony was inoculated in 5ml of LB containing 10µg/ml of rifampicin and cultured at 29°C for about 24 hour with vigorous shaking (240 rpm). The culture was then inoculated into 1L of 2YT and kept growing until the OD_{600} reached 0.8-1.0. The steps for cell harvesting and washing were exactly the same as that for *E coli*. The cells were finally resuspended in 2ml of ice-cold 10% glycerol, aliquoted, snap frozen and stored the same as for *E coli*.

For transformation, 30µl of competent cells was used to mix with 1µl (up to 0.5 µg) of binary plasmid prepared as in 2.1.1 and electroporated under the same settings as for *E coli*. After electroporation, 500 µl of SOC medium was added and the cells were allowed to recover for 2 hours at 29°C. 100 µl of cells was then spread on LB plate containing appropriate antibiotics and cultured at 29°C. Colonies normally appear after 2 days for kanamycin selection at 50 µg/ml or 3 days for chloramphenicol selection at 100 µg/ml. A single colony was picked, identified by PCR and propagated for plant transformation.

2.7.1.3 Transformation of Arabidopsis thaliana

Arabidopsis transformation was based on the vacuum infiltration method developed by Bechtold *et al.* (Bechtold *et al.*, 1993; Bechtold and Pelletier, 1998). This method was further simplified by Clough and Bent (Clough and Bent, 1998a), which eliminated the labor-intensive vacuum infiltration process and used only floral dip. The procedure described here is mainly based on this simplified method. *Agrobacterium tumefaciens* strain EHA105 harbouring an appropriate binary vector (mainly the constructs based on pCAMBIA1201) was used for transformation. To prepare for the inoculum (*Agrobacterium* suspension), small scale *Agrobacterium* cultures (3-5ml) were grown in liquid LB medium containing appropriate antibiotics (50µg/ml kanamycin or 100µg/ml chloramphenicol) at 28°C for overnight. The small scale cultures were then diluted 100-fold into LB medium with appropriate antibiotics and grown for roughly 18-24 hours. Cells were harvested by centrifugation for 10 min at 5000*g* and then resuspended in infiltration medium to a final OD_{600} of 0.8-1.0 prior to use.

Arabidopsis plants used for transformation were grown to flowering stage in 64 cm^2 pots (5-10 plants per pot) under the condition noted in 2.10.1. For floral dip, plants were inverted into a beaker containing the inoculum such that all above-ground tissues were submerged. After 5-10 sec, the dipped plants were removed from the beaker,

placed in a plastic tray and covered with Clingwrap. The plants were left in *Arabidopsis* growth room in normal condition for 24 hours before Clingwrap was removed. The plants were then allowed to grow to maturity and seeds were harvested. Seeds were dried at 37°C for three days prior to selection process (synchronized seed germination can usually be achieved by this treatment).

For selection of transformants, seeds were selected either on selection medium for hygromycin or kanamycin resistance (Clough and Bent, 1998a; Ye *et al.*, 1999), or on soil for the herbicide Basta resistance (Bechtold *et al.*, 1993; Mengiste *et al.*, 1997). For selection on medium, seeds were surface sterilized by first soaking in 10% bleach containing 0.05% Tween 20 for 5min and then in 70% ethanol for 10 min, followed by extensive washing with sterile water. The treated seeds were then distributed onto selection medium at a density of approximately 3000 seeds per 150×25mm^2 plate. Selection medium was germination medium (GM) plus 30μg/ml hygromycin or 50μg/ml kanamycin depending on the selectable marker used. Plates were sealed with surgical tape and placed in a culture room for seed germination (22 °C, 16-hour light/8-hour dark). After about 7-10 days, transformants, which can be identified by their bigger size over the non-transformed seedlings and normal root growth, were transplanted to soil for further growth and analysis.

For selection directly on soil, which is suitable for Basta selection, seeds were sown on soil in 30×40 cm trays and grown under the condition noted in 2.10.1. When the first pair of true leaves emerged for most of the seedlings (about a week), plants were sprayed with Basta (AgrEvo, 18% glufosinate) diluted to 100mg/L in water. Three such Basta spraying were applied in the interval of 5 days. True transformants could be easily identified by their normal appearance after the last spraying while the non-transformed plants bleached and died. The Basta-resistant plants were then transplanted to pots for further growth and analysis.

2.8 Histochemical localization of GUS activity

Arabidopsis tissues (inflorescence, leaf, root etc.) were examined for GUS (β-glucuronidase) activity essentially according to the procedure described by Jefferson *et al.* (Jefferson *et al.*, 1987; Jefferson, 1987) with some modification. The samples were immersed in GUS staining solution (50mM sodium phosphate pH 7.0, 10mM EDTA, 0.1% Triton X-100, 2mM potassium ferrocyanide, 2mM potassium ferricyanide, 100µg/ml chloramphenicol, 0.02% sodium azide, 1mg/ml X-Gluc), vacuum infiltrated for 5-10min and incubated at 37°C for overnight. The GUS stained tissues were then cleared in either a Clearing Solution (20% lactic acid, 20% glycerol, 1×PBS) (Vielle-Calzada *et al.*, 2000) or a modified Hoyer's solution (7.5g of gum arabic, 100 g of chloral hydrate, 5ml of glycerin and 30ml of H_2O) for at least 5 hours. The cleared tissues were observed directly using a Zeiss Axioskop or Leica DMR microscope. Cleared whole-mount ovules were observed under diferenctial interference contrast (Nomarski) optics. Images were recorded by a video/RGB camera attached to Zeiss Axioskop or a Nikon Coolpix 900 digital camera attached to Leica DMR.

For histochemical localization of GUS activity with high resolution, the GUS stained tissues were fixed in 3% glutaraldehyde, embedded in LRWhite Resin C° (Basingstoke, UK) and sectioned on a Reichert Ultracut Microtome by C. Miller in the Microscopy Center, CSIRO. Sections (2µm) were examined under Leica DMR microscope and images were taken as mentioned.

2.9 GFP detection

Flowers of *Arabidopsis* with enhanced Green Fluorescent Protein (EGFP) (Yang *et al.*, 1996) expression in ovules were collected and placed on a slide for dissection. To view ovules in a whole gynoecium (pistil), a gynoecium was removed from a flower and opened to reveal ovules using hypodermic needles (for 1ml insulin syringe) as scalpels.

Opened gynoecium was then transferred onto a plate with GM medium. EGFP was examined under a Leica MZ FL III fluorescence stereomicroscope with 470/40nm excitation filter and 525/50nm barrier filter. To view individual ovule with higher magnification, ovules were cut out from the carpels with two needles, transferred onto a slide with a drop of 1x PBS and then covered with a coverslip. EGFP expression in ovules were monitored with a Leica DMR microscope equipped with the same set of filters for GFP plant fluorescence. Images were taken either on a Nikon Coolpix 900 digital camera or using Fuji chrome 400F film.

2.10 Growth condition and cross of *Arabidopsis thaliana*

2.10.1 Growth condition

Arabidopsis thaliana ecotypes of Landsberg erecta, Columbia and C24 were used for the experiments. Seeds were germinated on soil containing one part of compost and one part of sand, and grown at 22°C with light set at 16-hour day (4000 lux) and 8-hour dark. 1×Hewitt nutrient solution was added to soil several times during the early plant growth stage.

2.10.2 *Arabidopsis* cross

Both pistil donor (female) and pollen donor (male) plants for cross were grown to flowering stage (4-6 weeks) under the condition noted in 2.10.1. Flower buds from pistil donor plants, which were at the right stage for crossing (short immature stamens with anthers that are greenish-yellow in color) were emasculated by removing stamens with forceps under a dissection microscope. After emasculation, the stigmas were pollinated with the pollen-releasing anthers from the pollen donor plants. To eliminate contaminating pollen, forceps were sterilized between crosses and handling male donors and female donors with 95% ethanol. When pollination was completed, the

pedicel of the pollinated flowers was painted with black ink as a mark and a labeled tag was attached to the plant. Following successful pollination, the pistil elongated to form normal silique. Seeds can be examined and harvested after about three weeks.

2.11 Media and buffers used in the experiments

BACTERIAL MEDIA

LB (Luria-Bertani) Medium (1 liter)

Bacto-tryptone	10g
Bacto-yeast extract	5g
NaCl	10g
Adjust pH to 7.0	

2YT Medium (1 liter)

Bacto-tryptone	16g
Bacto-yeast extract	10g
NaCl	5g
Adjust pH to 7.0	

SOC Medium (1 liter)

Bacto-tryptone	20g
Bacto-yeast extract	5g
NaCl	0.5g
$MgCl2{\cdot}6H_2O$	2g
KCl	0.2g
Glucose	3.6g

Adjust pH to 7.0

For solid medium, 15g/l of Bacto-agar was added.

Unless otherwise stated, all media were sterilized by autoclaving for 20 min at 121°C (15 lb/sq).

MEDIA FOR ARABIDOPSIS TRANSFORMATION AND SELECTION

Standard Infiltration medium (I liter)

0.5×MS Basal Medium (Sigma) (Murashige and Skoog, 1962)

1×B5 vitamins (Gamborg *et al.*, 1968)

5%	sucrose
44 nM	benzylaminopurine
0.5 g/L	MES

0.02%Silwet L-77 (Osi specialties, A Witco Company)

Adjust pH to 5.7

Simplified Infiltration medium (I liter)

5%	sucrose
0.05%	Silwet L-77

Germination medium (GM) (I liter)

0.5×	MS Basal Medium
1×	B5 vitamins
1%	sucrose
0.5 g/L	MES
0.5%	Phytagel (Sigma)

Adjust pH to 5.7

After autoclaving, add Timentin to a concentration of 100µg/ml (Dolferus *et al.*, 1994)

BUFFERS AND SOLUTIONS

TE (pH 8.0)

10 mM Tris·Cl (pH 8.0)

1 mM EDTA (pH8.0)

TAE (1×)

40 mM Tris-acetate

1 mM EDTA (pH 8.0)

TBE (0.5×)

45mM Tris-borate

1 mM EDTA (pH 8.0)

10×phosphate-buffered saline (PBS)

NaCl	8g
KCl	0.2g
Na_2HPO_4	1.44g
KH_2PO_4	0.24g

Adjust pH to 7.4

DNA Gel-loading Buffer (6×)

0.25% bromophenol blue

0.25% xylene cyanol FF

40% (w/v) sucrose in water

RNA Gel-loading Buffer

2× TAE (pH 8.3)

13% Ficoll

7 M urea

0.01% bromophenol

0.05% xylene cyanol FF

20× SSC (1 liter)

NaCl	175.3g
Sodium citrate	88.2g

Adjust pH to 7.0 with 10 N NaOH

5×HSB (1 liter)

100mM	PIPES
3M	NaCl
20mM	EDTA

Adjust pH to 6.8

10×Denhardt's III reagent (1 liter)

BSA fragment V	20g
SDS	100g
Ficoll 400	20g
PVP-360	20g

$Na_4P_2O_7 \cdot 10H_2O$	50g

5× Formaldehyde gel running buffer

0.1M	MOPS (pH 7.0)
40mM	sodium acetate
5mM	EDTA

Sterilized by filtration

10×Formaldehyde gel-loading buffer

50%	glycerol
1mM	EDTA
0.25%	bromophenol blue
0.25%	xylene cyanol FF

CHAPTER 3

ET253-AN ENHANCER TRAP LINE OF *ARABIDOPSIS THALIANA* WITH EGG APARATUS SPECIFIC GUS EXPRESSION PATTERN

3.1 Introduction

In *Arabidopsis thaliana*, large numbers of enhancer trap lines have been generated by several laboratories in the past several years (Sundaresan *et al.*, 1995; Klimyuk *et al.*, 1995; Campisi *et al.*, 1999; Parinov *et al.*, 1999). Among the enhancer trap lines examined by GUS staining, GUS expression in more than one tissue type seems to be dominant. This is possibly due to the non-specificity of the enhancers detected and that the associated genes may be turned on and off as required during development (Sundaresan *et al.*, 1995), or perhaps the multiple locations of the enhancer trap element DsE or T-DNA which reflect the effects from two or more enhancers with different functions (Campisi *et al.*, 1999).

As part of the efforts towards the identification of genes associated with apomixis, this project was initiated to isolate regulatory elements and/or their associated genes involved in the process of female gametophyte development by using particular enhancer trap line(s) with spatially restricted GUS expression patterns. The criteria for the screening of such an enhancer trap line may include: 1. exhibiting specific GUS expression in the early stages of megagametophyte development; 2. single copy insertion of the enhancer trap sequence (either DsE or T-DNA); 3. no obvious mutation phenotype and fertile. However, preliminary results have revealed that, in the DsE based enhancer trap system, approximately 2.7% of the enhancer trap lines showed GUS expression in the female gametophyte but most of these lines also

have GUS expression in other parts of the plant (U. Grossniklaus, unpublished data). Therefore, it is not easy to find a line that perfectly meets the above criteria. By examining around 2000 enhancer trap lines, Dr Grossniklaus' group was able to obtain some candidate lines that show GUS expression in different tissues (cells) in the female gametophyte such as the egg apparatus, oocyte, synergids and some in entire gametophyte. These enhancer trap lines could therefore be extremely useful in detecting genes expressed in specific female gametophyte cells (Grossniklaus *et al.*, 1995).

This chapter gives a short introductory description of the available Ds based enhancer trap lines, the selection of candidate enhancer trap lines with spatially restricted expression patterns and preliminary study on an enhancer trap line with specific GUS expression in megagametophyte cells.

3.2 Results

3.2.1 Enhancer trap lines with GUS expression in ovules

We obtained six primary Ds inserted enhancer trap lines of *Arabidopsis thaliana* from Dr Ueli Grossniklaus' lab. These were good candidates for further selection of line(s) with specific GUS expression in megagametophyte, especially in the egg cell. Morphologically, these lines are all normal in terms of both vegetative growth and fertility. They have been preliminarily examined by GUS staining in Grossniklaus' lab and the rough profiles of the GUS expression patterns were determined. As shown in Table 3.1, all these ET lines showed spatially restricted GUS expression in their ovules. Among these lines, ET133 and ET179 both revealed a complex GUS expression pattern with GUS staining not only in the floral tissues but also in the vegetative tissues. These patterns may reflect the influence from multiple cis elements and are difficult to analyze. They were therefore excluded for further examination.

For lines ET328, ET376 and ET456, they all have GUS expression in anthers and this put question mark on the tissue specificity of the possible enhancers detected in these lines, although the blue staining in anther is often an artifact in *Arabidopsis* (U. Grossniklaus, personal communication). Moreover, The GUS expression in line ET328 is mainly restricted in the synergid cells at early stages, suggesting that it is not related to the egg cell and the GUS staining in the whole female gametophyte is probably the result of synergid cell degeneration. In line 456, the GUS expression is in the egg apparatus but mainly occurs as a post-fertilization event. The line 376 shows quite general GUS expression in megagametophyte but its detailed profile of the GUS expression pattern still needs to be examined. All these lines could serve as good start materials for obtaining putative genes control certain steps in the reproductive process.

Table 3.1 Enhancer trap lines with GUS expression in the ovules (from Dr Ueli Grossniklaus' Lab)

Enhancer trap lines	GUS stained tissues		
	Ovule	Other floral tissues	Seedling
ET133	gametophyte & strip in chalazal region	anther	root, leaf, vasculature
ET179	egg apparatus, sometimes entire gametophyte	pollen, base of gynoecium	root, leaf, cotyledon
ET253	egg apparatus, sometimes entire gametophyte	no staining	no staining
ET328	synergids, sometimes entire gametophyte	pollen	no staining
ET376	pre-fertilization: egg apparatus; post-fertilization: entire gametophyte	anther	ND
ET456	egg apparatus, mainly post-fertilization	pollen	no staining

ND = not determined

The line ET 253 seems to be particularly suitable for the purpose of discovering female gametophyte specific genes and cis-elements, because it shows specific GUS expression in the embryo sac (mainly in egg apparatus) but no GUS expression in anthers and any other tissues of the plant. This means that the insertion site of the Ds

element in this line is most likely to be within a genomic region that controls the megagametophyte-specific gene expression. Therefore, this line was chosen for further study.

3.2.2 GUS expression profile at different floral stages in ET253

To examine the GUS expression in different stages of ET253 floral development, ET253 plants were grown under the conditions as noted in 2.10.1. After the plants reached the flowering stage, whole inflorescences with flowers/buds in different developmental stages and other vegetative tissues were collected and subjected to GUS staining as described in 2.8. For better observation of the gynoecium under the light microscope, the inflorescences with buds and flowers at different floral stages (pre- and post-fertilization) were placed on a slide after GUS staining and clearing. Half of the sepals, petals and stamens on each flower were removed to ensure the proper expose of the gynoecia. This arrangement made it easier to compare GUS expression patterns of flowers at different developmental stages.

The GUS staining results were consistent with Dr Grossniklaus' observation. GUS expression in ET253 is restricted in ovules and no GUS activity is detectable in leaf, stem, root and other floral tissues including stigma, carpel, sepal, petal and stamen. Furthermore, the results also revealed that the GUS expression is independent of fertilization but may be negatively regulated by fertilization. As shown in Figure 3.1, the GUS activity is not detectable before stage12. An example is flower 1, which is a small bud with gynoecium length around 0.5 mm and anthers still green. Flower 2, which is around stage 12, starts to show GUS activity in some of the ovules. With the development of ovule, the embryo sac elongates and the intensity of GUS staining increases. This covers the stages 12 and 13 as represented by flowers (buds) 3 and 4. During fertilization (stage14), the embryo sac becomes increasingly curved and the GUS activity remains high as shown in flower 5. The post-fertilization stages

(stages 15-16) show the decrease in GUS activity (flower 6) and no GUS expression can be observed in mature seeds. The floral stages used here refer to the table by Webb and Bowman (Mansfield and Bowman, 1994).

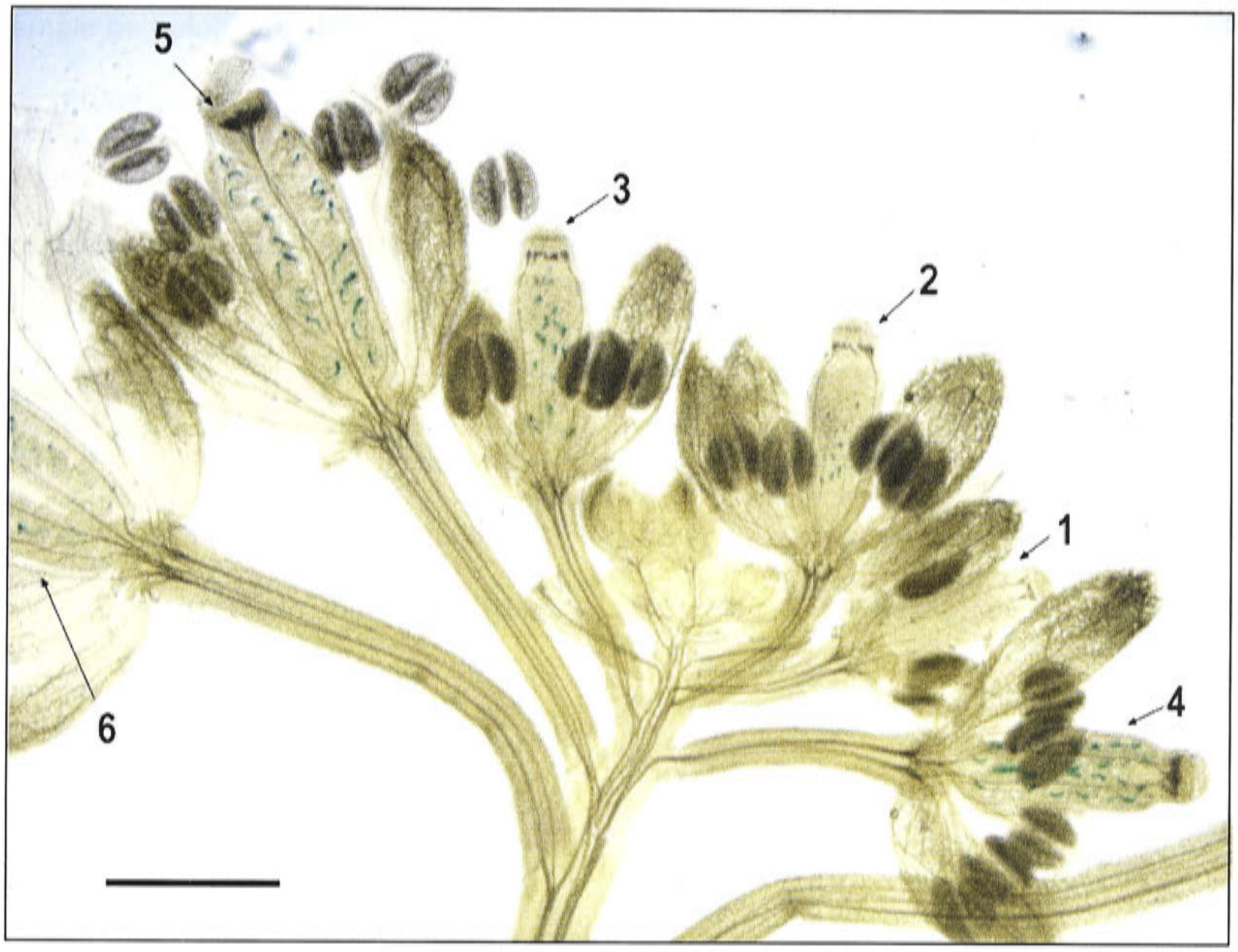


Figure 3.1 Megagametophyte-specific GUS expression of ET253 flowers at different developmental stages.

An inflorescence with flowers at different flowering stages from an ET253 plant was subjected to GUS staining. For photographing, half of sepals, petals and stamens from each numbered flower were removed to reveal the gynoecium. The approximate floral stages for the numbered flowers are: flower 1, stage 11; flower 2, stage mid 12; flower 3, stage late 12; flower 4, stage 13; flower 5, stage 14; flower 6, stage 15-16. Bar = 1 mm.

3.2.3 Further determination of GUS expression in the female gametophyte in ET253

To determine the ovule developmental stages which show GUS expression, some land markers associated with approximate floral stages (Mansfield and Bowman, 1994) and female gametophyte stages (Schneitz *et al.*, 1995; Christensen *et al.*, 1997) were used as references. GUS activity can be detected from flowers between stage 12 and stage15 in ET253. As shown in Figure 3.1, flower 1 is at floral stage 11, which does not show detectable GUS activity in the ovules (Figure 3.2-A). The GUS activity starts to be detectable from floral stage mid to late12 (flower 2, Figure 3.1), where the majority of ovules are at stage 3-IV (female gametophyte stage FG4). The ovules at this stage have the GUS activity restricted at the micropyle end and therefore could be from one of the micropylar nuclei (Figure 3.2-B). The clear boundary between the stained and non-stained areas in the embryo sac is due to the existence of the large central vacuole.

After cellularization, the embryo sac elongates to form the seven-celled female gametophyte (ovule stages 3-V to 3-VI or female gametophyte stages FG5 to FG6) and the intensity of GUS staining increases but the GUS staining is mainly located at the egg apparatus (Figure 3.2-C). This was confirmed by sectioning of the ovule at this stage (Figure 3.2-F). This covers the floral stages 12 and 13 (flowers 3 and 4, Figure 3.1). After fertilization (stage14, flower 5), the zygote still shows strong GUS activity (Figure 3.2-D) but the GUS activity decreases afterwards and no GUS expression can be observed in ovules after mid globular embryo stage (Figure 3.2-E). This observation suggested that the GUS expression in ET253 is probably restricted to the early stages of female gametophyte development. Pollination may have a negative effect on the GUS expression.

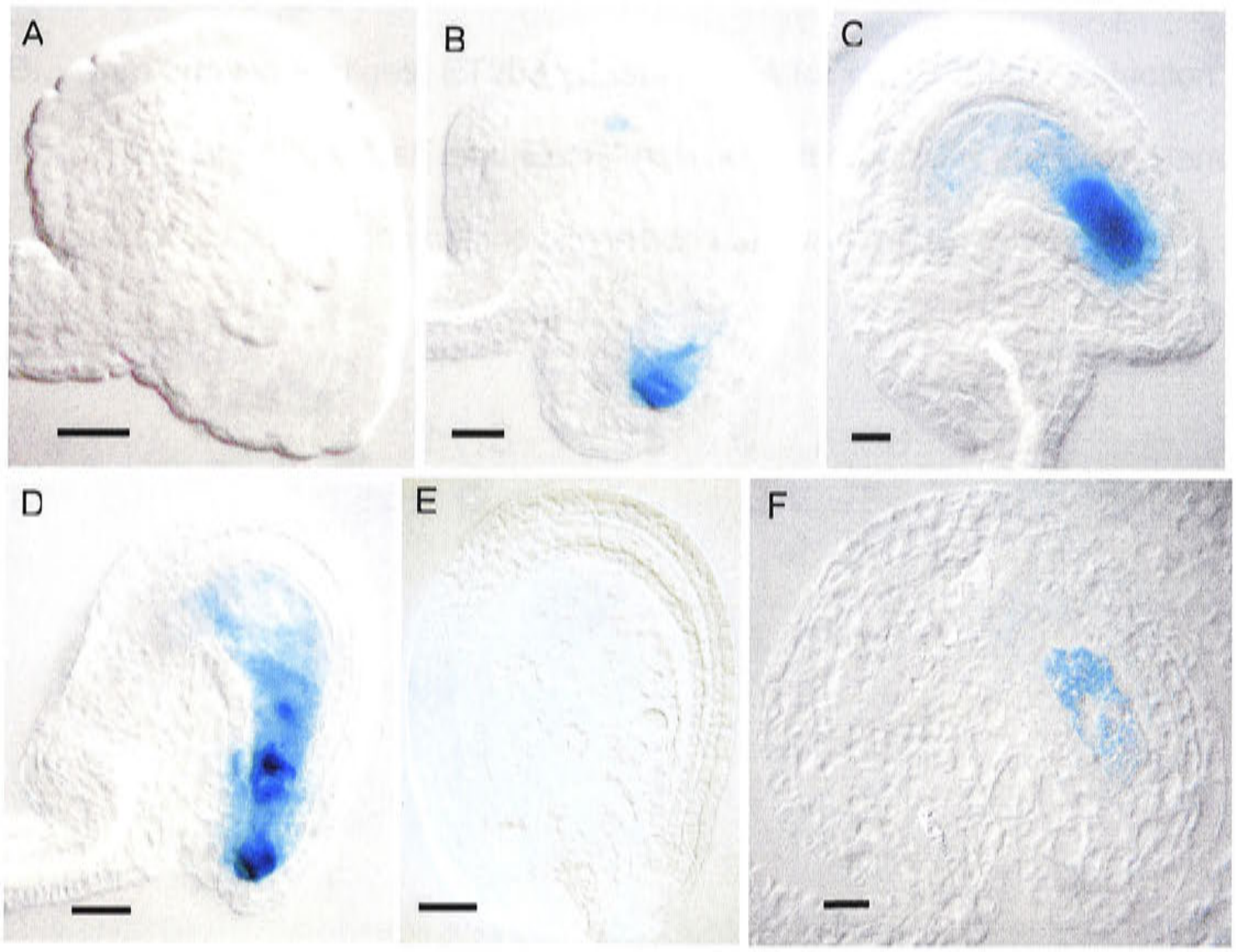


Figure 3.2 GUS expression at different ovule developmental stages in ET253.

(A-E) Optical sections of X-gluc stained and cleared whole-mount ovules at different stages. After GUS staining, Ovules from flowers at different stages were cleared by modified Hoyer's solution (refer to 2.8, Chapter 2) and observed under Nomarski optics.

(A) an ovule from flower 1 (ovule stage 3-I to 3-II). (B) an ovule from flower 2 (ovule stage 3-IV). (C) an ovule from flower 3 or 4 (ovule stages 3-V to 3-VI). (D) an ovule from flower 5 (ovule stage 4-IV to 4-V). (E) immature seed with embryo at mid globular stage from flower 6 (stage after 4-VI).

(F) Longitudinal section (2μm) of a GUS stained ovule at around the same stage as in (C).

Bars in (A) through (C) and (F) = 10 μm; Bar in (D) =20 μm; Bar in (E) = 50

3.2.4 ET253 contains only a single DsE insertion

Although it is believed that most of the Ds insertion lines contain a single Ds element (Sundaresan *et al.*, 1995; Parinov *et al.*, 1999), this still needs to be confirmed for line ET253. Southern analysis was carried out to determine the copy number of the DsE insertion in ET253. Genomic DNA from ET253 and wild type Landsberg erecta (Ler)

was prepared as described in 2.1.2. Three restriction enzymes, HindIII, EcoRI and SnaBI, were chosen to digest ET253 genomic DNA for copy number detection. HindIII has no cutting site in the DsE and EcoRI has no cutting site in the *gusA* gene in the DsE, whereas SnaBI has one cutting site inside the *gusA* gene.

Figure 3.3 Confirmation of single DsE insertion in ET253 genome by Southern blot analysis.

Approximately 3 μg of genomic DNA from both ET253 and Landsberg erecta (Ler, negative control) were digested with EcoRI, HindIII and SnaBI, respectively, separated on 0.7% agarose gel and transferred onto positively charged nylon membrane. A 0.96 kb α-^{32}P-dATP labeled MfeI/BstBI fragment from the *gusA* gene was used as probe in hybridization as described in 2.5.1. Lane 1, λ/BstEII DNA markers; lane 2, ET253/HindIII; lane 3, Ler/HindIII; lane 4, ET253/EcoRI; lane 5, Ler/EcoRI; lane 6, ET253/SnaBI; lane 7, Ler/SnaBI; lane 8, λ/HindIII DNA markers.

When the 0.96 kb MfeI/BstBI fragment from the *gusA* gene is used as probe, the number of signals detected in HindIII and EcoRI digested genomic DNA should represent the copy number of the DsE inserts but SnaBI digested DNA should show

two signals for each DsE insert. Figure 3.3 showed the hybridization results. It can be seen that only one hybridization band was detected from either HindIII (lane 2) or EcoRI digested genomic DNA (lane 4), while two close bands were detected from SnaBI digested genomic DNA (lane 6), confirming that the DsE insertion in ET253 is single copy. The corresponding negative controls from Ler (Lane 3, 5 and 7) showed no hybridization signal, indicating that there is no *gusA* homologous sequence in *Arabidopsis* genome.

3.3 Discussion

3.3.1 Enhancer trap: transposon based system and T-DNA based system

In *Arabidopsis*, enhancer trap lines have been generated by means of either transposable elements (Sundaresan *et al.*, 1995; Klimyuk *et al.*, 1995) or direct T-DNA insertions (Campisi *et al.*, 1999). The approach using direct T-DNA insertion is technically easier than the approach based on the two component Ac/Ds system, with the advent of T-DNA transformation by vacuum infiltration (Bechtold *et al.*, 1993) or floral dip (Clough and Bent, 1998). As a result, it is relatively easy to generate large numbers of random enhancer trap insertions in the genome. Both approaches have produced plenty of enhancer trap lines with the reporter gene expression pattern of interest in different tissues or organs (Martienssen, 1998; Campisi *et al.*, 1999). However, although large collections of T-DNA transformants for enhancer trap have been produced, analyzed and scores of interesting patterns identified, it has not been easy to make the exact correlation between the reporter gene expression pattern and the corresponding gene or the regulatory element. This is mainly due to multiple T-DNA insertions in different locations in the genome, thereby complicating interpretation of expression patterns. In contrast, the insertions generated by Ds transposition are mostly single copy of the intact element, reducing the possibility of artifactual patterns

of reporter gene expression induced from multiple tandem or rearranged insertions. I therefore chose to work with the Ds based enhancer trap lines exhibiting certain GUS expression in the embryo sac.

3.3.2 GUS staining in anthers of some enhancer trap lines

In *Arabidopsis*, GUS staining in anther and/or pollen is often observed when GUS expression is used for testing promoter function. In the case of enhancer trap lines examined, many also displayed blue color to certain degree in anthers or pollen. This is often attributed as an artifact in many occasions (U. Grossniklaus, personal communication). To confirm this phenomenon, I conducted GUS staining using wild type plants from ecotypes Landsberg erecta and C24 under the same staining condition as that for transgenic plants (normally staining at 37°C for up to 24 hours). Results showed that all the vegetative tissues had no GUS staining but anthers from some plants of both Landsberg and 24 exhibit weak blue color. However, this mainly happens in the anthers of young buds before floral stage 12. The young buds with blue staining were further examined by sectioning and observation under dark field microscopy (Klimyuk *et al.*, 1995). Typical orange-coloured indigo blue crystals, which are the product of X-gluc after GUS cleavage, can be clearly seen in the anthers with blue staining when transgenic plants with GUS expression was compared under dark field (data not shown). The reason for the deposit of the orange crystals in some young anthers of wild type plants is yet unknown.

From Table 3.1, we can see that all the listed lines, except for line ET253, show blue staining in anthers. This could be due to the following possibilities: (1) artifact, (2) enhancer or cis-element in a promoter detected which controls gene expression in both male and female parts, and (3) more than one enhancer or cis-element in a promoter detected which control gene expression in male and female parts separately. If it is the first case, then lines 328 and 456 could have GUS expression specifically in the female

gametophyte. Cis-elements responsible for specific gene expression in the synergid cells or the egg apparatus could be identified from these lines. If the lines showing GUS staining in both male and female parts belong to the second case, cis-elements and genes that play a role in both male and female gametophytes (as described in 1.2.3.4) could then be uncovered. If the scenario falls into the last case, the identification process could be complicated but this should be the rare case.

3.3.3 Determination of the floral and ovule developmental stages showing GUS expression in ET 253

In ET253, we used some standards such as the pistil (gynoecium) length and the appearance of flower outlined by Webb and Bowman (1994) to estimate the floral stages. Since the standards used by Webb and Bowman are mainly based on ecotype Columbia while the line ET 253 is with the background of ecotype Landsberg erecta, small adjustment to the standards was needed to have a better estimation. For example, the pistil length of Landsberg erecta is generally slightly shorter than that of Columbia when the growth condition is the same. However, the appearance of the flower is very much the same for the two ecotypes. Stamens and petals are the two best references for certain floral developmental stages. A flower (bud) that shows green anthers without visible filaments is generally at a floral stage prior to stage 11 when meiosis is yet to occur. During megasporogenesis (floral stage 11), the anthers are still green but the filaments elongate while the petals are also green. The petals turning to white while levelling with the long stamens and the anthers turning to yellow mark the end of megasporogenesis and the start of megagametogenesis (floral stage 12). In Figure 3.1, we removed petals from flowers 1 to 4 to well expose the gynoecia because these flowers were still buds at the time of examination and therefore could not show the relative lengths of petals in these flowers. However, the length of filaments is also a good indicator of floral stages and they were well shown in Figure

3.1. By using these obvious land markers, I am able to determine the floral stages of each GUS stained flower to a very close approximation.

The ovule developmental stages with observable GUS expression were determined generally based on the staging system established by Schneitz *et al.* (1995). This is the most detailed staging system available at present for determining ovule developmental stages in *Arabidopsis*. Although a very good staging system for female gametophyte development was also established by Christensen *et al.* (1997), I could not apply it alone to determine the female gametophyte stages because this system is based on the observation using confocal laser scanning microscopy, which is not suitable for the observation of GUS stained tissues.

3.3.4 Better reporter gene may be needed for enhancer trap system in detecting expression patterns restricted to only a few cells

In determining the stages of GUS stained ovules, I had some problems in keeping the balance between the intensity of GUS expression and the structure maintaining of the embryo sac. For normal GUS staining, tissues need to be incubated in GUS staining solution (2.8, Chapter 2) at 37°C for at least 24 hours to ensure the detection of low level GUS expression. However, long incubation in GUS staining solution makes the observation of the fine structure of embryo sac more difficult due to the delicacy nature of the ovule tissues. As a result, incubation time has to be reduced as a compromise for maintaining better ovule structure. Meanwhile, I found that incubation at 30°C did not significantly affect the GUS staining level but could still keep good ovule structure. Therefore, I routinely carry out GUS staining of ovules at 30°C for 16 hours before clearing and mounting. Since this is a compromised way of detecting GUS expression and considering the sensitivity of the *E. coli* GUS, some important information on certain tissue/cell types may still be missing when GUS expression is weak. For instance, weak GUS expression in megaspore mother cell may not be detected. In the

future, a more sensitive reporter gene may need to be used for this purpose. A good candidate is the GUS Plus reporter gene system recently developed by CAMBIA (www.cambia.org). GUS Plus was isolated from bacterium *Staphylococcus* sp. The gene has been codon optimized for high expression in plants and has better catalytic activity for more rapid detection of GUS activity (T. Nguyen and R. A. Jefferson, unpublished data). Test of this gene in rice and *Arabidopsis* showed that it is more sensitive and needs less staining time (T. Nguyen, unpublished data). This gene could therefore be a better choice to be used as a reporter gene in gene trap enhancer or trap for detection of female gametophyte specific genes and enhancers.

3.3.5 Possibilities for the enhancer(s) detected by the enhancer trap in ET 253

The purpose of my research was to identify and isolate genes and/or cis-regulatory elements trapped by DsE from the genomic regions flanking the insertions using the enhancer trap lines displaying certain GUS expression in the embryo sac. In ET 253, the GUS expression pattern is unique. This certainly reduced the complexity in identifying the causal agent if the lines with more complex GUS expression patterns are compared. However, there are still three possibilities underlying this megagametophyte-specific expression pattern. First, the GUS expression pattern could be controlled by a cis-element within the promoter of a nearby gene. This is the best of the three possibilities because it may lead to the identification of not only the cis-element itself but also its associated gene. This gene then could most probably be a gene that plays an important role in the megagametophyte development. Second, the GUS expression pattern could reflect the effect from a distal enhancer element, which interacts with a promoter of certain gene through long-range activation. Some examples of such action exist in organisms with large genomes such as *Drosophila* and metazoans. In these genomes, regulatory elements that determine the patterns of gene expression frequently act over thousands of base pairs (Dorsett, 1999; Mahmoudi

et al., 2002). For instance, the *Drosophila* homeotic *Ultrabithorax* (*Ubx*) gene has many enhancers and silencers spread out over 100 kb that regulate the gene function (Farkas *et al.*, 2000). One key regulatory sequence, a Polycomb response element (PRE), is located about 25 kb upstream of the *Ubx* transcription start site (Pirrotta, 1997; Lyko and Paro, 1999). Moreover, some evidences showed that gene can even be activated *in trans* through the mediation of the transcription factor GAGA (Mahmoudi *et al.*, 2002). This complex long-range effect is difficult to identify if it happens in enhancer trap lines. Hopefully, the small size of the *Arabidopsis* genome will not have too many genes in such arrangement. Third, the GUS expression pattern could be caused by a cryptic promoter. A cryptic promoter leading to the seed coat-specific GUS expression has been reported in tobacco (Fobert *et al.*, 1994) and another cryptic constitutive promoter sequence, tCUP, was also reported (Foster *et al.*, 1999). These cryptic promoters were detected by T-DNA tagging using promoter trap constructs. This could occur in enhancer detection system as well to pick up spurious enhancers that are not associated with the expression of the adjacent genes.

The confirmation of one Ds insertion in line ET253 ensures that the identification of the regulatory element for the megagametophyte-specific gene expression is straightforward. Since line ET253 is morphologically and genetically normal, I speculate that the DsE insertion in the genome of this line may not be in a gene or the insertion could also be in a redundant gene, which will not lead to any obvious mutant phenotype.

CHAPTER 4

CHARACTERIZATION OF GENOMIC REGION ASSOCIATED WITH MEGAGAMETOPHYTE-SPECIFIC EXPRESSION

4.1 Introduction

As a first step towards the identification of any cis-elements or genes tagged by transposon or T-DNA insertions, genomic DNA sequences flanking the insertion must be recovered. To meet the need for such a purpose, several PCR-based techniques have been developed by using the known DNA sequence within the transposon or T-DNA. These include inverse PCR (IPCR) (Ochman *et al.*, 1990; Ochman *et al.*, 1993), thermal asymmetric interlaced PCR (TAIL-PCR) (Liu *et al.*, 1995) and transposon insertion display (TID) (Yephremov and Saedler, 2000). In principle, all the PCR methods aimed at the amplification of DNA fragments outside the boundaries of known sequences need to construct a template for the PCR first. This can be achieved either before PCR or in the PCR processes. To construct template before PCR, either circularization of DNA fragments containing known sequence as in IPCR (Ochman *et al.*, 1993) or ligation of linkers to digested genomic DNA as in TID (Yephremov and Saedler, 2000) were used. To generate template in the PCR processes, degenerate primers, when paired with arbitrary primers, were designed and used for loosely annealing to flanking region of known sequence (Liu *et al.*, 1995). Among these methods, TAIL-PCR has proven to be very effective in obtaining flanking sequences from Ds insertion lines (Parinov *et al.*, 1999) and T-DNA insertion lines (Campisi *et al.*, 1999). Comparing to TAIL-PCR, IPCR is technically more difficult but it is the method of choice for obtaining longer flanking sequences from both ends of the known sequence by one PCR reaction. TID is mainly designed for three classes of maize transposons *Cin4*, *mu1* and *En/Spm*, and therefore was not chosen for this study.

Once enough DsE flanking sequence is obtained, detailed analysis of the sequence can be carried out in several ways. First, candidate genes or expressed sequence tags (ESTs) (Adams *et al.*, 1991) could be found by BLAST sequence similarity searching against all known sequences in GenBank or other related databases. Second, the candidate genes can be examined for specific expression patterns by Northern analysis or *in situ* hybridization. Finally, the responsible sequence should be tested *in vivo* by transforming back into plant, in order to confirm its function.

This chapter presents the results for the acquisition of the DsE flanking sequences from the enhancer trap line ET253 by TAIL-PCR and IPCR, and some further analyses of this region for candidate genes responsible for the megagametophyte-specific expression pattern.

4.2 Results

4.2.1 Amplification of the direct DsE flanking region in ET253 by TAIL-PCR

The isolation work was initiated by attempting to amplify both the 5' and 3' region flanking the Ds element in ET 253 by TAIL-PCR. As shown in Figure 4.1, three specific nested primers Ds5-1, Ds5-2 and Ds5-3 for the amplification of 5' flanking fragments and three specific nested primers Ds3-1, Ds3-2 and Ds3-3 for the amplification of 3' flanking fragments were designed and synthesized according to the orientation of the Ds element. However, since the *gusA* reporter gene was placed in the opposite direction to Ds sequence in the DsE (Figure 4.1A), the orientation of the DsE flanking region in ET253 was then determined in the opposite direction accordingly. The DsE flanking regions upstream and downstream of the *gusA* gene were therefore designated as 5' and 3' regions, respectively. This change makes it easier to present data for the follow up analysis when the sequences to be tested for enhancer function are fused to *gusA* gene. Two arbitrary degenerate primers AD1 and AD2, which are

exactly the same as described by Liu *et al.* (1995), were also synthesized. The sequences of these primers were listed in Table 4.1.

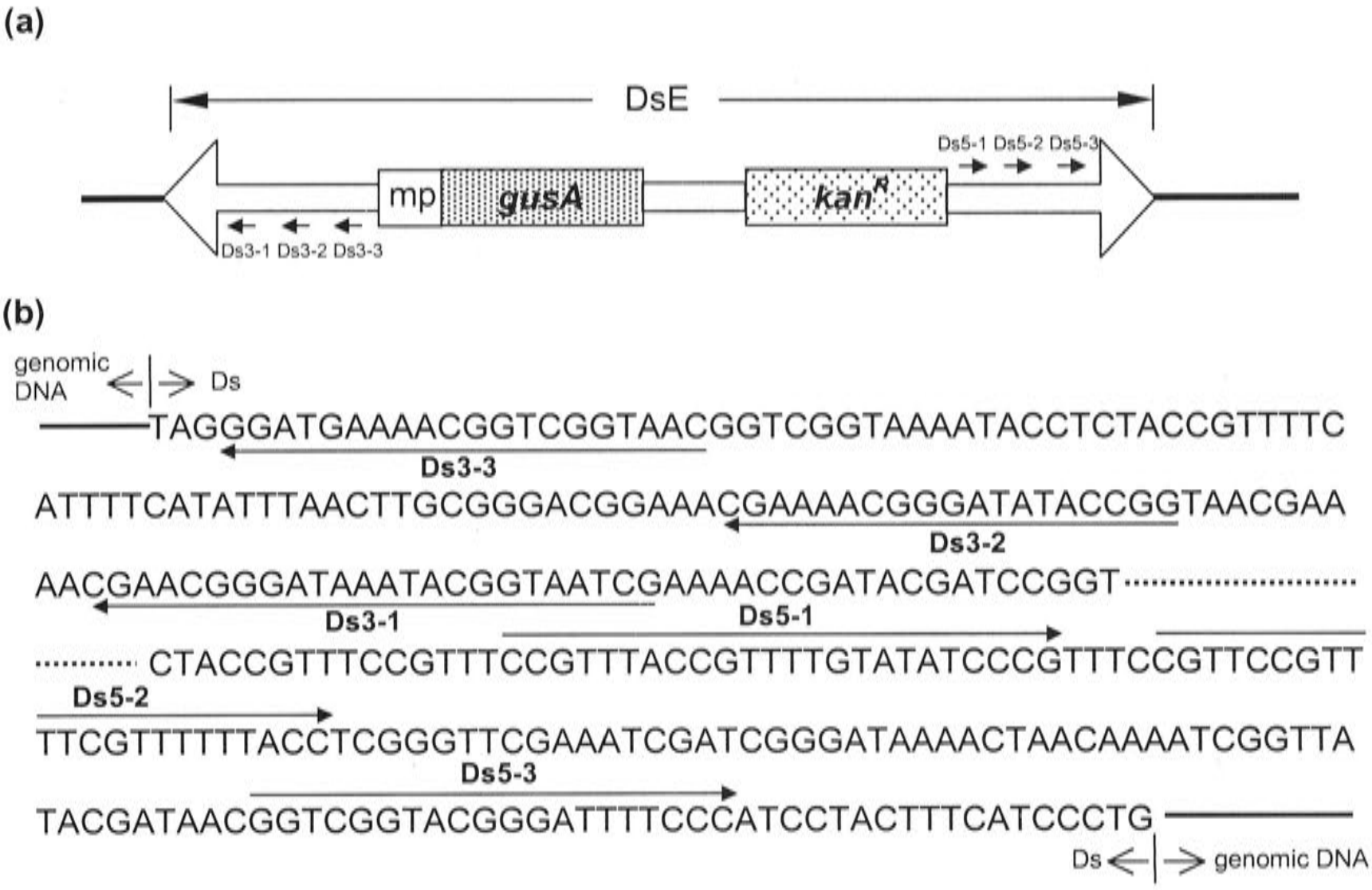


Figure 4.1 Design of Ds-specific nested primers for TAIL-PCR.

(a) Schematic diagram of the enhancer trap element in Ds (DsE) (Sundaresan *et al.*, 1995) indicating the position of the specific primers at both the 5' and 3' ends. The orientation of the DsE follows the orientation of the GUS reporter gene (*gusA*). The Ds element is depicted as open two-headed arrow and the flanking genomic regions are denoted as thick lines. Nested primers are shown as small arrows. *gusA*, β-glucuronidase gene; *kan*^{R}, kanamycin resistance gene; mp, CaMV 35S minimal promoter (-1 to -46 region).

(b) Detailed location of the nested specific primers Ds5-1, Ds5-2, Ds5-3 and Ds3-1, Ds3-2, Ds3-3 in DsE. The directions of the primers are indicated by arrows. The flanking genomic regions are shown as solid lines and the middle part (including the *gusA* and *kan*^{R}) in the DsE was simplified as dotted line.

Table 4.1 Nucleotide sequence of Ds-specific primers and arbitrary degenerate (AD) primers.

Ds5-1	CCG TTT ACC GTT TTG TAT ATC CCG	24 mer
Ds5-2	CGT TCC GTT TTC GTT TTT TAC C	22mer
Ds5-3	GGT CGG TAC GGA ATT CTC CC	20 mer
Ds3-1	CGA TTA CCG TAT TTA TCC CGT TCG	24 mer
Ds3-2	CCG GTA TAT CCC GTT TTC G	17 mer
Ds3-3	GTT ACC GAC CGT TTT CAT CC	20 mer
AD1[a]	NTC GA(G/C) T(A/T)T (G/C)G(A/T) GTT	15 mer
AD2[b]	NGT CGA (G/C)(A/T)G ANA (A/T)GA A	16 mer

[a] 64-fold degeneracy, average Tm = 45.3°C

[b] 128-fold degeneracy, average Tm = 46.6°C

The secondary TAIL-PCR products for both 5'and 3' flanking sequences were obtained from Dr Ueli Grossniklaus and used as tertiary TAIL-PCR templates. With the attempt to amplify specific fragments from both 5' and 3' regions, primer combinations of Ds5-3/AD1, Ds3-3/AD1, Ds5-3/AD2, Ds3-3/AD2 were tried in tertiary PCR. An approximately 170 bp tertiary PCR fragment was obtained from 5' secondary TAIL-PCR products by Ds3-3/AD2 combination, while other primer combinations failed to generate any specific fragment (Figure 4.2). After sequencing, the 170 bp PCR fragment was confirmed to be the correct product because it contains the Ds sequence. About 150 bp 5' DsE flanking genomic sequence was obtained (Figure 4.3). This sequence was then used as a start point for further inverse PCR amplification of more DsE flanking region in ET253.

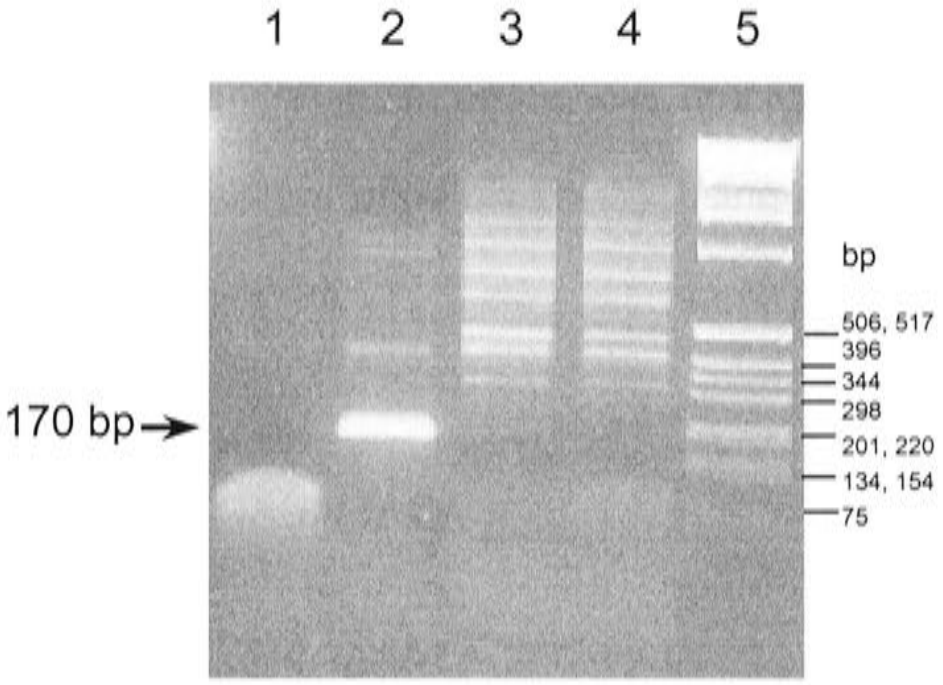


Figure 4.2 Tertiary TAIL-PCR using the secondary TAIL-PCR products from ET253.

Primary and Secondary TAIL-PCR were performed by Dr U. Grossniklaus using genomic DNA from line ET 253 as template and the primer combinations of Ds5-1/AD1, Ds3-1/AD1, Ds5-1/AD2, Ds3-1/AD2 (primary) and Ds5-2/AD1, Ds3-2/AD1, Ds5-2/AD2, Ds3-2/AD2 (secondary). The tertiary PCR was performed using the 1250-fold diluted secondary PCR products as templates. The tertiary TAIL-PCR products were run on 1.6% agarose gel in 1×TAE. The primer combinations are: lane 1, Ds5-3/AD2; lane 2, Ds3-3/AD2; lane 3, Ds5-3/AD1; lane 4, Ds3-3/AD1. Lane 5 is the 1kb ladder DNA markers. An approximately 170 bp specific band (indicated by arrow) was obtained from the primer combination of Ds3-3/AD2 (lane 2), while no specific products were obtained from other primer combinations.

4.2.2 Further isolation and sequencing of the DsE flanking region

Since an enhancer can often work efficiently at quite a distance from the transcription start site either upstream or downstream (Maniatis *et al.*, 1987; Blackwood and Kadonaga, 1998), long sequences flanking the DsE need to be obtained from both the upstream and downstream regions. Although it is not guaranteed to include the putative magagemetophyte-specific enhancer sequence(s) within certain distance of the DsE flanking region as discussed in Chapter 3, the chance of getting the responsible enhancer element(s) would be higher if longer DsE flanking sequence can

be available and tested. In this case, the limitations of TAIL-PCR make it no longer the preferred technique. Two techniques can be used to obtain longer flanking region containing the putative enhancer sequence: amplification of the genomic DNA by inverse PCR or screening a genomic library using the small sequence identified in TAIL-PCR as probe.

ET-62T:	CCA CGA TGC AAA TAT ATC GAT AAC G	25 mer
ET-121B:	ATT AAT CTT GGG GTA ACT TTA CTT C	25 mer

Figure 4.3 153bp sequence from the 170 bp tertiary TAIL-PCR fragment and primers designed for the first inverse PCR.

The 170 bp tertiary TAIL-PCR fragment was sequenced using primer Ds3-3 and confirmed to be the DsE flanking sequence. The 153 bp sequence shown here is of the complementary strand to the sequenced strand. Two oligonucleotide primers suitable for inverse PCR were designed based on this sequence.

The inverse PCR method was first tried and shown to be very effective in obtaining sequences from both 5' and 3' flanking regions of a known sequence. For our purpose, two specific primers (ET-62T and ET-121B) were designed, based on the 150 bp 5' DsE flanking sequence obtained by TAIL-PCR (Figure 4.3), and used in the subsequent inverse PCR reaction using genomic DNA from *Arabidopsis thaliana* ecotype C24 as template. Three different restriction enzyme digestions (BamHI, EcoRI, and HindIII) were tried. A 4.8 kb specific fragment was amplified from the template

DNA that had been digested by HindIII and circularized, while no amplification was obtained from the template DNA that had been digested by BamHI or EcoRI before circularization (Figure 4.4). The reason for using C24 DNA was that the genomic DNA from line ET253 was not available at the time when the project was initiated and it was assumed that the sequence around the DsE in ET253 (derivative of ecotype *Landsberg erecta*) should be the same as that in C24. By sequencing this 4.8 kb inverse PCR fragment, enough DsE flanking sequence data, which covers 3.6 kb upstream region and 1.2 kb downstream region, was obtained that made it possible to design primers for further inverse PCR to obtain more DsE flanking sequence.

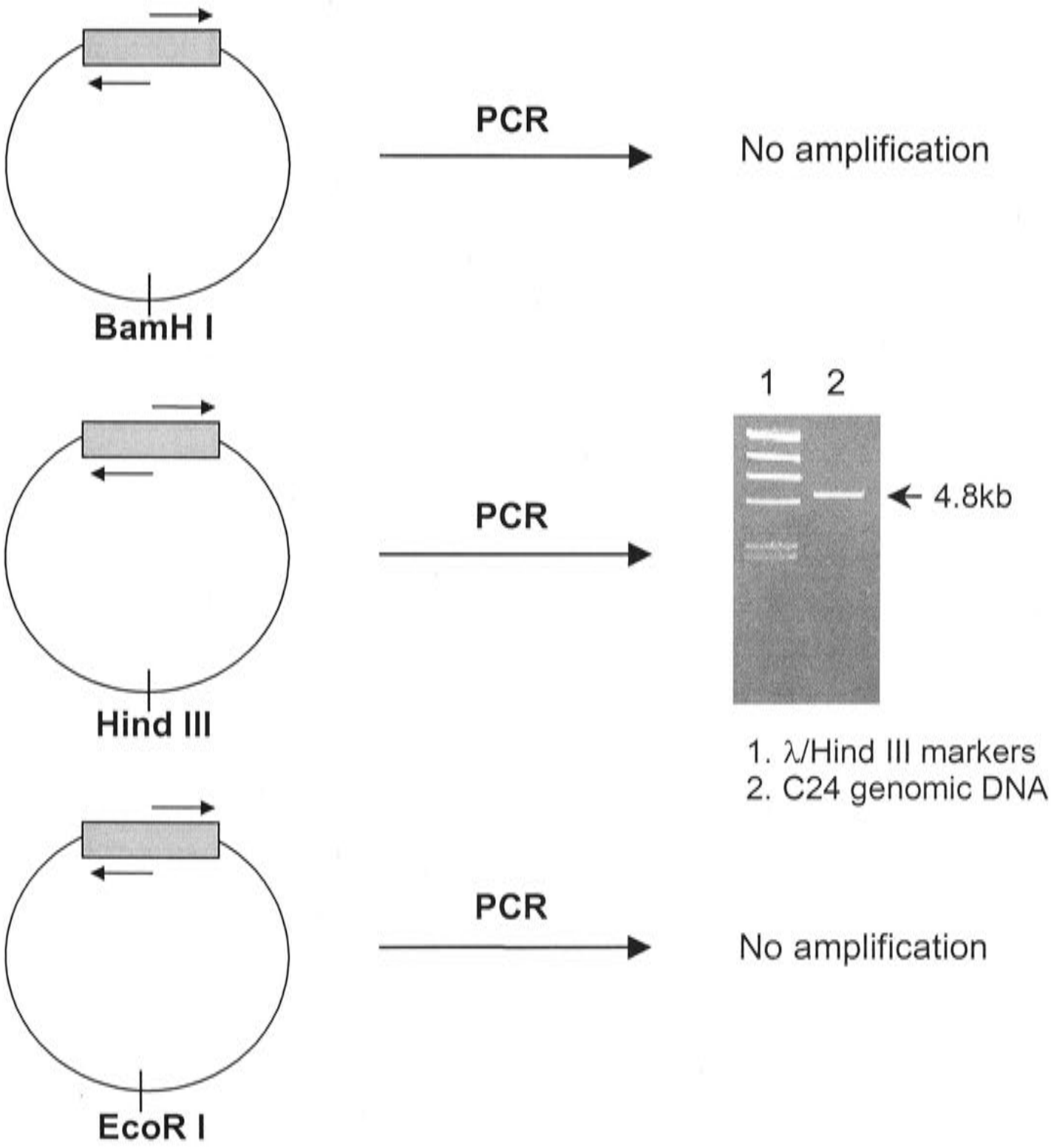


Figure 4.4 First inverse PCR using primers designed from the 150 bp DsE flanking sequence derived from TAIL-PCR.

Genomic DNA from ecotype C24 was digested with BamHI, HindIII and EcoRI, seperatelyly and circularized to form inverse PCR templates. PCR was carried out using primers ET-62T and ET-121B. Specific amplification only occurred in PCR reaction involving HindIII treated genomic DNA. This specific PCR product was estimated 4.8 Kb in size. The gray box and its surrounding circle represent the 150 bp known sequence from TAIL-PCR and its flanking region, respectively. The small arrows above and under the gray box represent primers ET-62T and ET-121B, respectively.

After genomic DNA from ET253 became available, the further inverse PCR was carried out using ET253 DNA as template. As shown in Figure 4.5, two more rounds of inverse PCR were designed to obtain more 5' and 3' flanking sequence using EcoRI and SnaBI digestion for circular template preparation. By testing different combinations of primers designed based on the *gusA* gene and the 4.8 kb DsE flanking sequence from C24, the pairs ET-1411B/GUSA+1403 and ET+730T/GUSA-648 were shown to be good for the specific amplification of the further 5' and 3' region, respectively. 6.3 kb new sequence of 5' region was then obtained from PCR using EcoRI derived circular template with primers ET-1411B/GUSA+1403, and 1.8 kb new sequence of 3' region was obtained from PCR using SnaBI derived circular template with primers ET+730T/GUSA-648. A total of 8.5 kb 5' and 3 kb 3' DsE flanking sequences were obtained from ET253. All the 11.5 kb DsE flanking region was sequenced (See Appendix B). The sequencing result showed that there is a significant difference between the genomic sequence from ET 253 and the corresponding sequence from C24, beyond the nucleotide -1737 upstream of the DsE (the DsE insertion site was designated as 0) (data not shown). This sudden change of sequence in C24 made it impossible to continue to use C24 sequence for further identification of the sequence responsible for the megagametophyte-specific GUS expression. Therefore, only the sequence data from ET253 was used in the subsequent research.

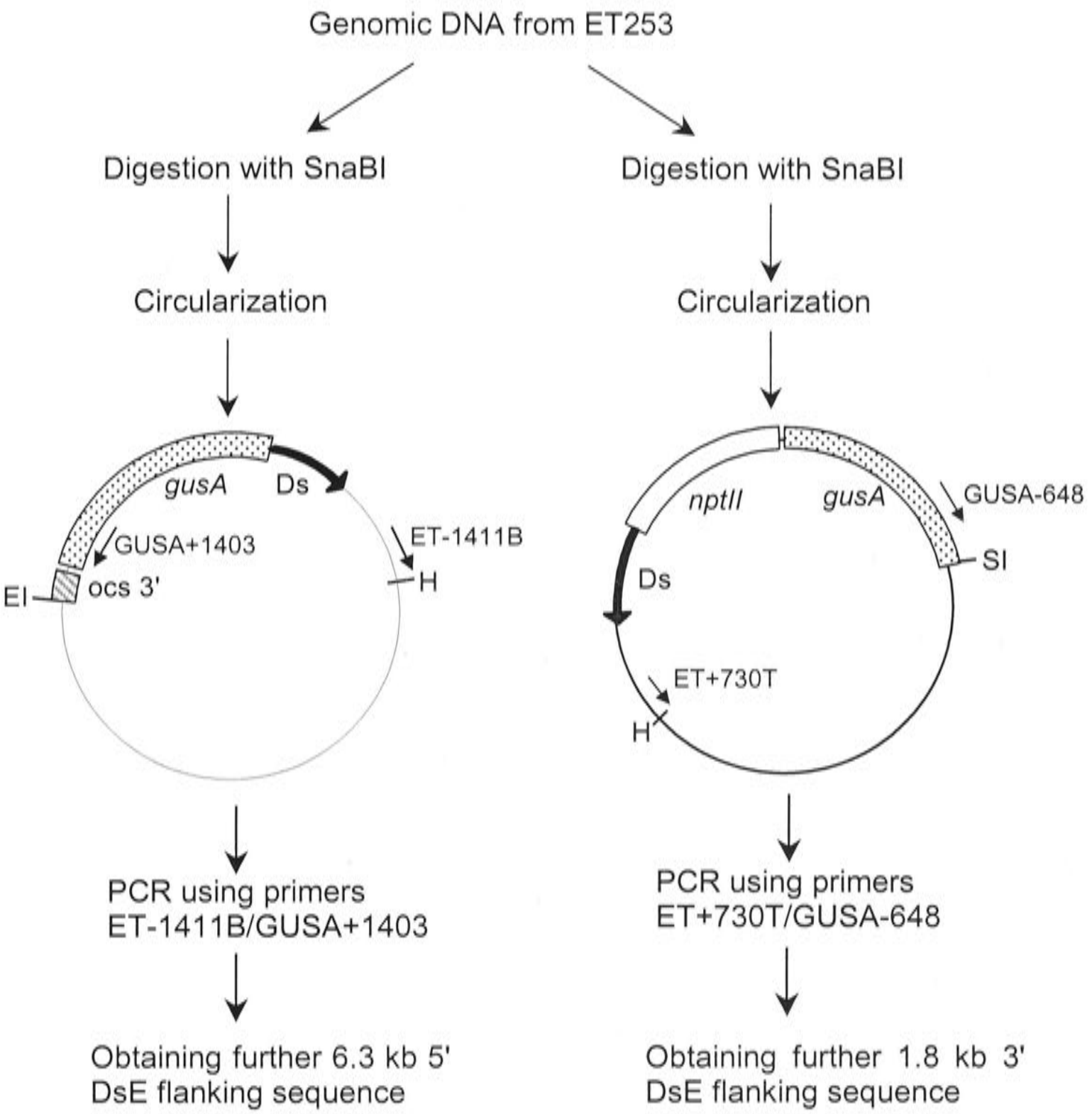


Figure 4.5 Further inverse PCR amplification of the DsE flanking region in ET253. EcoRI (EI) and SnaBI (SI) were used to digest genomic DNA from ET253. The digested DNA fragments were circularized to form templates for PCR amplification of the sequence in 5' and 3' DsE flanking region beyond the HindIII (H) site. Primers are indicated by small arrows and their sequences are as follows: ET-1411B, 5'-ACTGTGGCTCTGTCTCTGTG-3'; ET+730T, 5'-TCG-TACTAATGTCACCTGGA-3'; GUSA-648, 5'-CCACCTGCCAGTCAACAGAC-3'; GUSA+1403, 5'-ATGTTCTGCGACGCTCACAC-3'.

4.2.3 BLAST search of the 11.5kb DsE flanking sequence

When the 11.5kb sequence of *Arabidopsis* genomic DNA from enhancer trap line ET253 was obtained in early 1998, the *Arabidopsis* sequence information in GenBank was still quite limited. Now, the whole *Arabidopsis* genome sequence has been completed using the ecotype Columbia (Mayer *et al.*, 1999; Walbot, 2000). Sequence alignment of the 11.5 kb sequence we have obtained from line ET253 showed that this 11.5kb sequence (from ecotype Landsberg erecta) is 98% identical to the corresponding sequence from ecotype Columbia. It is therefore possible to examine the flanking region of the DsE insertion site in ET253 to spot some candidate genes, which may be associated with certain megagametophyte-specific enhancer (MGSE) element.

Similarity search using basic local alignment search tool (BLAST) (Altschul *et al.*, 1997) was conducted in GenBank to find the candidate genes. The DsE insertion site in ET253 is located at the top arm of the chromosome 4. It is known that insertion of Ds element results in formation of transposon footprint, which is a duplication of 8 bp sequence at the insertion site (pohlman *et al.*, 1984; Scott *et al.*, 1996). The 8bp duplication sequence in ET253 was identified as ATTAAGGC by comparing the sequences around the DsE insertion site between ET253 and wild type Landsberg erecta (Figure 4.6a). The BLAST searching revealed several genes near the DsE insertion site. The upstream region has two genes (named here as *atpH21-1* and *atpH21-2*), 2.1 kb and 5.2 kb away, respectively (Figure 4.6b). The two genes are highly homologous to a peroxidase gene *atp9a* (GenBank accession No: X98856) (Justesen *et al.*, 1998) or otherwise named as prxr2 gene (GenBank accession No: X98314). The downstream region has a gene coding for an unknown predicted-protein that matches with an EST tag (GenBank accession No: R65200), 1.4 kb away (Figure 4.6b), and a gene coding for formamidase, 5.5 kb away (referred to BAC clone F19F18, GenBank accession No: AL035605, data not shown). The two peroxidase genes

atpH21-1 and *atpH21-2* share almost the same exon regions but are different in their introns. All the above genes can be good candidates for the megagametophyte-specific expression pattern.

(a)

.....TATGAT**ATTAAGGC** taqqqa-***DsE***-tccctq **ATTAAGGC**TCCGA .....

(b)

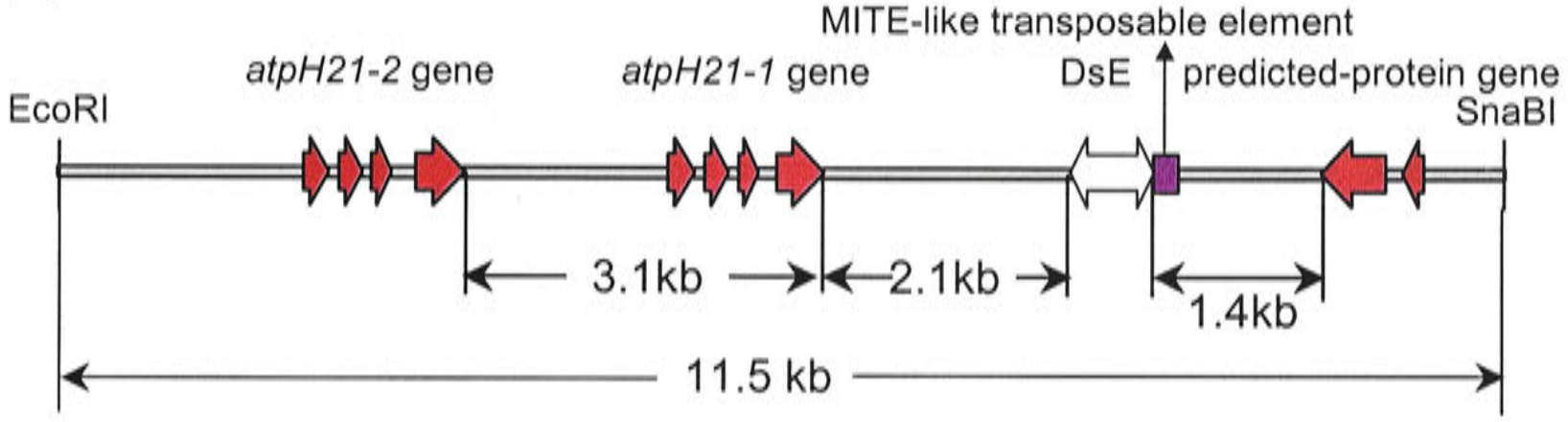


Figure 4.6 The genomic region on chromosom 4 covering the DsE insertion site in ET253

(a) The sequence around the DsE insertion site in ET253. The duplication of 8 bp sequence after DsE insertion was confirmed and shown in bold type (underlined).

(b) BLAST searching revealed two peroxidase genes *atpH21-1* and *atpH21-2*, 2.1 kb and 5.2 kb apart from the DsE insertion site, respectively, in the upstream region, and a predicted protein gene, 1.4 kb downstream of the DsE insertion site. The exons and the directions are shown in red arrows. The inserted DsE is denoted as open left-right arrow and the MITE-like transposable element is shown as purple box.

Interestingly, the 165 bp sequence directly downstream of the DsE insertion site (after the 8bp DsE duplication sequence ATTAAGGC) in ET253 is highly homologous to many such sequences in five chromosomes in different locations. More than 30 such sequences can be revealed by BLAST search. Twelve clones containing this repetitive sequence were selected to represent the distribution of this sequence in all five

chromosomes (Figure 4.7a). These include bacterial artificial chromosome (BAC) clones (F19F18, F11F19, F28P10, F13O11, T26J13, T22J18, F9H16, T15B16 and F2H17), phage (P1) clones (MWI23 and MQK4) and transformation-competent artificial chromosome (TAC) clone (K6M13).

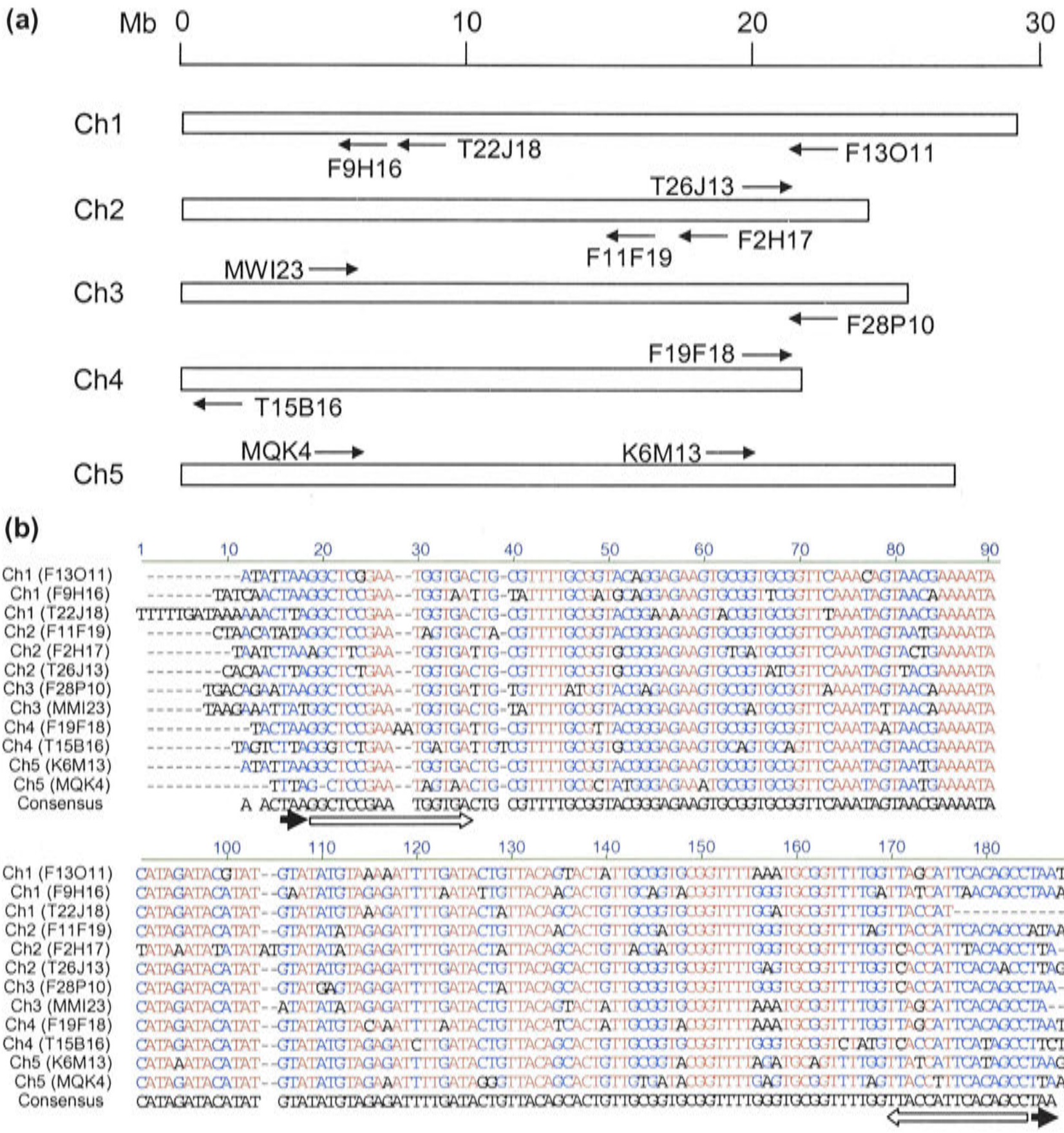


Figure 4.7 The MITE-like sequence directly downstream of the DsE insertion site in ET253 is widely spread in *Arabidopsis* genome.

(a) Distribution of the representative MITE-like sequences on five chromosomes. Each chromosome is depicted as an open box. Twelve BAC clones containing the MITE-like sequences are selected from AtDB's *Arabidopsis* Sequence Map Overview to represent the existence of the MITE-like sequences on all five

chromosomes. Chromosomes 1 to 5 are abbreviated as Ch1 to Ch5. The chromosomal locations of relevant BAC clones are indicated by arrows (not in scale).

(b) Alignment of the twelve selected MITE-like sequences. Nucleotides identical in all twelve sequences are in red color and those identical in most of the sequences are in blue color. The terminal inverted repeat sequences are indicated by open arrows and the target site duplications (TAA) are indicated by black arrows. The corresponding BAC clone names are in parentheses.

Alignment of the twelve selected sequences showed that they are highly conserved (Figure 4.7b). These small repetitive sequences are AT-rich, have imperfect terminal inverted repeats (TIRs), and do not seem to have any coding capacity. In addition, the majority of such sequences are flanked by the TA(A) target site duplication (Figure 4.7b). All these features suggest that these repetitive sequences can be a new member of the miniature inverted-repeat transposable elements (MITEs) family (Wessler *et al.*, 1995; Casacuberta *et al.*, 1998). Therefore, the MGSE element detected by the enhancer trap is also possibly associated with this MITE-like transposable element.

4.2.4 Southern and Northern blotting analysis of the two peroxidase genes upstream of the DsE insertion site

Southern analysis was first applied to determine the copy number of the two peroxidase genes *atpH21-1* and *atpH21-2*. These two peroxidase genes are highly homologous to the peroxidase gene *atp9a* (Justesen *et al.*, 1998). Genomic DNA was prepared as described in 2.1.2 from wild type Landsberg erecta and digested with BglII, AccI and NcoI plus NheI, respectively. After blotting onto nylon membrane, the restricted genomic DNA was hybridized using an α-^{32}P-dATP labeled 540bp SpeI/DraI fragment from the 4th exon of the *atpH21-1* gene as probe (Figure 4.8a).

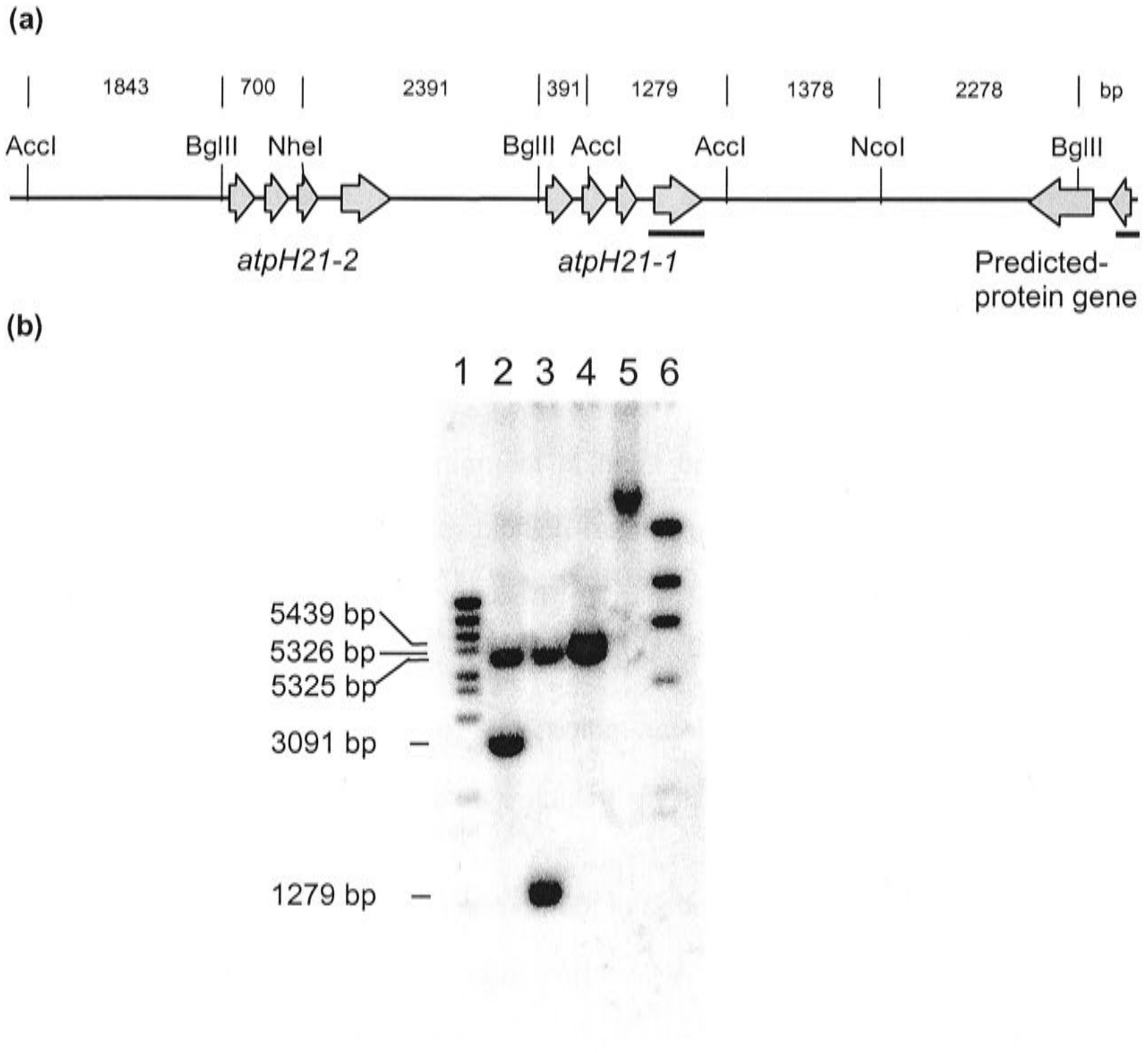


Figure 4.8 Southern analysis of the peroxidase genes *atpH21-1* and *atpH21-2*: copy number determination.

(a) Location of the restriction enzyme cutting sites chosen for Southern analysis. The numbers on top are the distance between two sites (in base pairs). The two bars under the 4th exon of the *atpH21-1* gene and the 1st exon of the predicted-protein gene indicate the regions used as probes for hybridization.

(b) Southern blotting. Genomic DNA from Landsberg erecta (Ler) was digested with BglII, AccI and NcoI plus NheI, respectively. The digested DNA was then separated on 0.8% agarose gel, blotted onto positively charged nylon membrane and probed with α-32P-dATP labeled 540bp SpeI/DraI fragment from the 4th exon of the *atpH21-1* gene. lane 1, λ/BstEII DNA markers; lane 2, Ler DNA/BglII; lane 3, Ler DNA/AccI; lane 4, Ler DNA/NcoI+NheI; lane 5, Ler DNA/uncut; lane 6,

λ/HindIII DNA markers. The sizes of the expected fragments from different digestions were shown on the left.

As shown in Figure 4.8a, BglII separates *atpH21-1* and *atpH21-2* into two fragments with one fragment containing *atpH21-1* at 5326 bp and the other fragment containing *atpH21-2* at 3091 bp. AccI cutting also generates two fragments when the 4th exon is targeted, but the sizes of the two fragments change to 5325 bp and 1279 bp. When the genomic DNA is digested by NcoI together with NheI, the two peroxidase genes should present as one fragment at 5439 bp if it is probed with the 540 bp SpeI/DraI fragment. The Southern blotting result clearly showed the above expected bands and no extra band was detected from each corresponding enzyme digestion (Figure 4.8b). Some weak signals on the blot should be non-specific due to the lack of high stringency wash. Therefore, the region harboring the two peroxidase genes *atpH21-1* and *atpH21-2* is present as a single copy in the genome.

Northern blotting was then carried out to determine if the peroxidase genes are expressed in a tissue-specific manner. Total RNA and mRNA were prepared from different tissues of Landsberg erecta plant: gyneocium, other floral tissues (mainly petal and sepal) and leaf. The RNA samples were then run on 1% denaturing agarose gel and blotted onto nylon membrane as described in 2.5.2. The membrane was then hybridized with the same probe as for the Southern blotting experiment described above (the 540 bp SpeI/DraI fragment). The Northern blotting result is shown in Figure 4.9. It can be seen from this picture that all tissues (gyneocium, other floral tissues and leaf) exhibited a band with the size of 1.2 kilo-bases, which is the expected size for the mRNA of the peroxidase genes *atpH21-1* and *atpH21-2*. Although it has been revealed from Southern blotting that the proxidase genes *atpH21-1* and *atpH21-2* are unique in the genome, their broad pattern of expression raised question about their roles in the female gametophyte development.

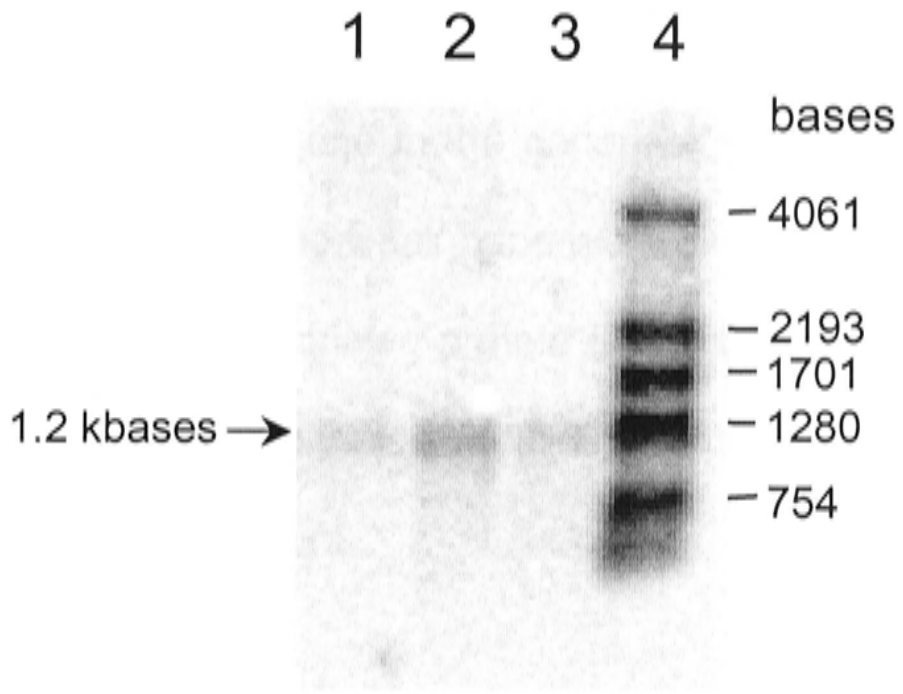


Figure 4.9 Northern blotting analysis of the peroxidase gene: expression in different tissues

mRNA were extracted from gynoecium, other floral tissues (mainly petal and sepal) and leaf of Ler ecotype. Approximately 0.5ug of mRNA for each sample was separated on 1% denaturing agarose gel, blotted onto positively charged nylon membrane and probed with the same α-^{32}P-dATP labeled 540bp Spe I/DraI fragment from the 4th exon of the atpH21-1 gene. lane 1, mRNA from gynoecium; lane 2, mRNA from other floral tissues; lane 3, mRNA from leaf; lane 4, RNA markers. The sizes of the RNA markers are shown on the right side. The hybridized signal is approximately 1.2 kilobases in size.

4.2.5 Southern and Northern blotting analysis of the predicted-protein gene downstream of the DsE insertion site

Southern analysis was also carried out to determine the copy number of the predicted-protein gene downstream of the DsE insertion site. Since the restriction sites BglII, AccI, NcoI and NheI can also be used for copy number determination of the predicted-protein gene, the membrane used for probing the peroxidase genes (in 4.2.4) was also used in this second experiment by reprobing it with a new probe. The 540 bp SpeI/DraI probe was stripped from the membrane, and the membrane was then reprobed with an α-^{32}P-dATP labeled 255 bp fragment from the putative first exon of the predicted-protein gene (Figure 4.8a). The 255 bp fragment matches with the EST tag (GenBank accession No: R65200) and is therefore within a transcribed region. All the enzymes (BglII, AccI, NcoI and NheI) used for digestion of genomic DNA do not cut inside the

255 bp sequence and therefore each hybridization signal in a single digestion should represent one copy of the gene in the genome. The reprobed blot showed that three strong bands and at least one weak band were detected (Figure 4.10). It suggests that the gene coding for the predicted-protein is a multiple copy gene. The hybridization signals vary among different bands, indicating that the gene sequences of the different copies may be slightly different.

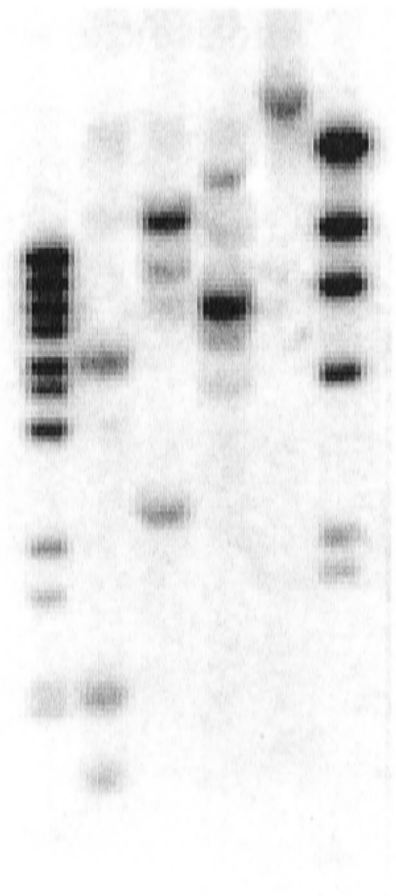


Figure 4.10 Southern analysis of the predicted-protein gene: copy number determination.
The same membrane for Southern blotting of the peroxidase genes (Figure 4.8) was reused here. After stripping off the 540bp Spe I/DraI probe, the membrane was reprobed with the α-^{32}P-dATP labeled 255bp EST matching fragment from the second exon of the predicted-protein gene. lane 1, λ/BstEII DNA markers; lane 2, Ler DNA/BglII; lane 3, Ler DNA/AccI; lane 4, Ler DNA/NcoI+NheI; lane 5, Ler DNA/uncut; lane 6, λ/HindIII DNA markers.

To detect the specificity of the predicted-protein gene expression, Northern blotting of mRNA from different tissues of Landsberg erecta was also performed using

the same membrane as in 4.2.4. The membrane was first treated to remove the 540 bp SpeI/DraI probe and then reprobed with the 255bp probe (the same as for Southern blotting). Figure 4.11 shows the result from this Northern hybridization. It was expected that multiple signals might be obtained but in fact only one band was detected from each tissue type. Obviously, the band from gynoecium and other floral tissues shifted to approximately 1 kilo-bases in size, which is consistent with the size of the mRNA for the predicted-protein gene, but the band from leaf is about 1,1 kilo-bases. The mRNA exists mainly in floral tissues (gynoecium and other floral tissues) and less in leaf tissue, compared to the peroxidase gene expression.

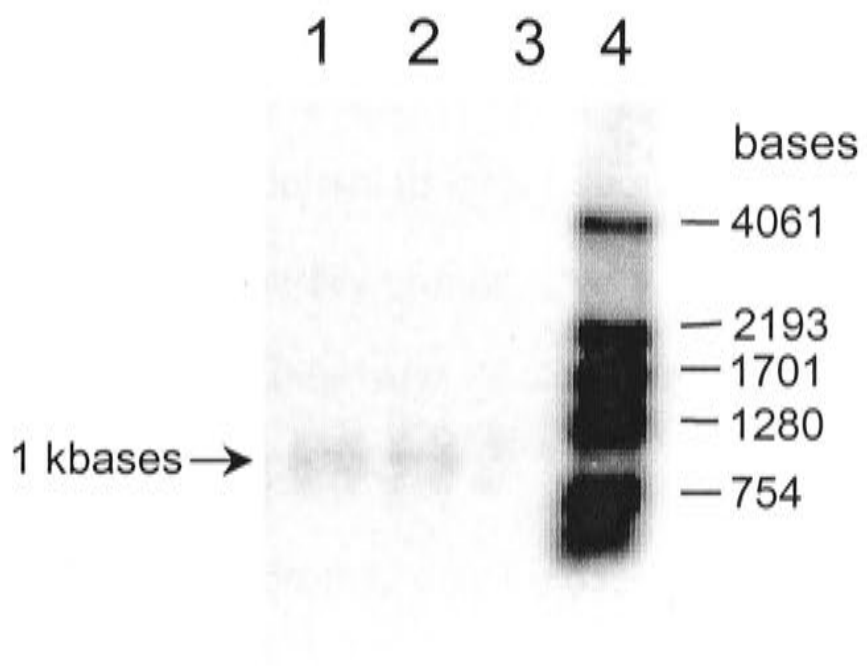


Figure 4.11 Nothern blotting analysis of the predicted protein gene expression. The same membrane for Northern blotting of the peroxidase genes (Figure 4.9) was reused here. After stripping off the 540bp Spe I/DraI probe, the membrane was reprobed with the same α-^{32}P-dATP labeled 255bp EST matching fragment from the second exon of the predicted protein gene as for Southern blotting. Lane 1, mRNA from gynoecium; lane 2, mRNA from other floral tissues; lane 3, mRNA from leaf; lane 4, RNA markers. The sizes of the RNA markers are shown on the right side. The hybridized signal is approximately 1 kbases in size.

More BLAST searching also showed that the putative exons of the predicted-protein gene were almost identical to an mRNA obtained from *Arabidopsis* roots

(GenBank accession: AV552189)(Asamizu *et al.*, 2000) and also to the mRNA for ASYMMETRIC LEAVES2-like protein 41 (GenBank accession: AB080843) (Iwakawa *et al.*, 2002). These results indicate that the predicted-protein gene is widely expressed in *Arabidopsis* plants. It therefore strongly suggests that the predicted-protein gene is less likely to be associated with the MGSE.

4.3 Discussion

4.3.1 The combination of TAIL-PCR and IPCR: a good way for obtaining large flanking DNA sequences

It has been shown in 4.2.1 and 4.2.2 that the combination of TAIL-PCR and inverse PCR gave good results in obtaining larger sequence flanking a known region. In TAIL-PCR, the choice of the arbitrary primer is very important. In this study, AD2 (128-fold degeneracy) seems better than AD1 (64-fold degeneracy), probably due to the better flexibility that AD2 possesses (Liu *et al.*, 1995). The size of the TAIL-PCR products is not crucial because it will normally only be used as a sequence tag for the insertion site of a certain Ds element or T-DNA. Once a specific TAIL-PCR product is obtained and sequenced, inverse PCR or screening of DNA library can then be applied based on the sequence information from the TAIL-PCR tag. Direct inverse PCR using the same enzyme digestions (HindIII, BamHI and EcoRI) and specific primers designed from Ds sequence was also tried but no result was obtained. One reason is the existence of these sites inside the DsE that prevent the formation of circular template with the whole DsE sequence. Another possible reason is the big size of DsE element (more than 6 kb) adding to the fragment, which is more difficult to be circularized than smaller DNA fragment to form good template for PCR. Therefore, the combination of TAIL-PCR and IPCR provides a good solution.

In IPCR, the selection of restriction enzymes for the digestion of genomic DNA is an important step that determines the size of the IPCR product. It is normally preferable to obtain larger IPCR fragments for each IPCR reaction for chromosome walking (Ochman *et al.*, 1993) so that more sequence can be obtained with less IPCR reactions. However, it seems that the IPCR product has certain size limit due to the difficulty in circular template formation when larger IPCR products were attempted. For instance, the EcoRI fragment covering the 5' flanking region of the DsE insertion site in ET253 is large in size (over 10 kb, as shown by the Southern analysis in 3.2.4), and it is possibly the main reason for the failure in obtaining IPCR product when EcoRI was used for genomic DNA digestion (as shown in Figure 4.4). The failure with BamHI digestion is probably due to the same reason.

4.3.2 Comparison of the DsE flanking regions among ecotypes Landsberg erecta, C24 and Columbia

The 11.5kb DsE flanking sequence that we determined comes from ecotype Landsberg erecta, while the publicly available complete sequence of the *Arabidopsis* genome was determined in ecotype Columbia (Col-0). Comparative analysis carried out by the Arabidopsis Genome Initiative using 82 Mb of the genome sequence of Columbia and 92.1Mb of the available sequence of Landsberg erecta revealed two classes of differences between the sequences: single nucleotide polymorphisms (SNPs) and insertion-deletions (The *Arabidopsis* Genome Initiative, 2000). In this study, sequence alignment between the two ecotypes in the 11.5 kb region revealed that the two ecotypes share 98% identity. Single-nucleotide-polymorphism (SNP) obviously present but no insertion or deletion was detected in this region when only Landsberg erecta and Columbia were compared (data not shown). However, as mentioned in 4.2.2, sequence comparison also revealed that there is a sudden sequence change in ecotype C24 beyond the nucleotide -1737 upstream of the DsE when the ET253

sequence (from Landsberg erecta background) was compared if the DsE insertion position is designated as 0. This could be due to an insertion, a deletion or a rearrangement in C24 genome. Since even SNPs could sometimes be very important in determining gene patterns and functions (Brown, 2000), only DsE flanking sequences from the original ET253 was used in the subsequent reconstruction and transformation experiments for the identification of the megagametophyte-specific enhancer to avoid any confusion.

4.3.3 Candidate genes and transposable element possibly associated with the megagametopyte-specific enhancer

BLAST search of the *A. thaliana* genome revealed only three genes in this 11.5 kb DsE flanking region in ET253: two *atp9a*-like peroxidase genes as tandem arrays and a predicted-protein gene (unknown gene with EST match). This is consistent with the general gene density on chromosome 4, which is 4.6 kb per gene (The *Arabidopsis* Genome Initiative, 2000). As shown in Figure 4.6b, the DsE landed in the middle of the region between the *atpH21-1* gene and the predicted-protein gene in ET253. According to the orientation of these two genes, the DsE insertion is in the downstream region of both genes. The limited space between the *atpH21-1* gene and the predicted-protein gene, which is only 3.4 kb between the stop codons of the two genes, and the lack of an unambiguous open reading frame (ORF), suggest that the 3.4 kb sequence where the DsE inserted may be an intergenic region. This partially explains why ET253 shows no mutant phenotype, although displaying mutant phenotype is a rare event in enhancer trap lines.

According to the arrangement of the 11.5 kb region flanking the DsE in ET253, one will naturally regard the three nearby genes *atpH21-1*, *atpH21-2* and the predicted-protein gene as candidates responsible for the megagametophyte-specific GUS expression pattern. As discussed in 3.3.5, several possibilities exist. The first possibility

here could be that a megagametophyte-specific cis-element in the promoter of one of the above genes can influence the GUS expression pattern. In this case, the gene harboring this cis-element could be a gene involved in the process of embryo sac development. However, the identification of the gene is now complicated by the fact that the three nearby genes *atpH21-1*, *atpH21-2* and the predicted-protein gene are not simply single copy gene. The two peroxidase genes *atpH21-1* and *atpH21-2* are very similarly arranged while the predicted-protein gene exists as multiple copies in the genome. It is therefore impossible to determine the function of each individual gene before we pinpoint the cis-element. The second possibility could be that an enhancer responsible for megagametophyte-specific expression could specifically interact with the promoter of one of the above genes but is located far from the gene it acts on. The third possibility could be that the megagametophyte-specific enhancer is not associated with one of three genes but is associated with a gene further away. If this is true, it is then necessary to look at other genes beyond the above three genes. In addition, the possibility could also be that the megagametophyte-specific enhancer is part of the MITE-like transposible elements, as described in 4.2.3. In this case, the enhancer should present as multiple copies in the genome.

4.3.4 The upstream peroxidase genes *atpH21-1* and *atpH21-2*

The peroxidase family is a superfamily consisting of three classes of peroxidases according to their origin: bacterial, mitochondrial and chloroplast (class I), fungal (class II), and classical plant (class III) peroxidases (Kjaersgard *et al.*, 1997). Plant peroxidases play a wide variety of functional roles related to defense, development, lignification, and hormonal signaling (Ostergaard *et al.*, 1998). Some peroxidases are expressed constitutively (Lavid *et al.*, 2001), whereas some others are inducible in response to external stress, such as wounding (Kawaoka *et al.*, 1994), pathogen infection (Mohan *et al.*, 1993; Curtis *et al.*, 1997) and salt stress (Botella *et al.*, 1994).

Although most peroxidases are expressed in a non tissue specific manner, organ- or tissue-specific peroxidases, such as root-specific (Wanapu and Shinmyo, 1996), stem-specific (Omann *et al.*, 1994) and endosperm-specific (Rasmussen *et al.*, 1991) peroxidases were also reported. In *Arabidopsis*, more than 40 peroxidase genes are now known but functional association is complicated by a general lack of peroxidase substrate specificity (Ostergaard *et al.*, 1998).

The two peroxidase genes *atpH21-1* and *atpH21-2* are located upstream of the *gusA* gene in the DsE of the enhancer trap line ET253 and they are single copy genes as identified by Southern analysis (4.2.4). These facts make them good candidates for the possible role as female gametophyte specific gene. Alignment of the sequences of the two genes revealed that their putative exons are almost the same but their putative introns are quite different. In addition, the putative promoter regions of the two genes also share some common sequences with overall 84% identical nucleotides within 1.2 kb. This kind of arrangement makes it possible that *atpH21-1* and *atpH21-2* encode the same proxidase but they may be expressed in different tissues and involved in different pathways in development. This is at least supported by the Northern analysis, which showed only one band with the same size from all tissue types tested (gyneocium, other floral tissues and leaf) (Figure 4.9). It is then possible that *atpH21-1* is specifically expressed during female gametophyte development while *atpH21-2* is involved more generally in the vegetative growth. This speculation means that the two peroxidase genes could play different roles in different pathways.

4.3.5 The downstream predicted-protein gene

The predicted-protein gene has been shown to be a multiple copy gene as determined by Southern analysis (Figure 4.10). A latest BLAST search showed that the predicted-protein gene downstream of the DsE insertion site in ET253 has a perfect match with a newly identified gene coding for ASYMMETRIC LEAVES2-like protein 41 (GenBank

accession: AB080843). The mRNA obtained from *Arabidopsis* roots (GenBank accession: AV552189) could be the product of the same gene. The ASYMMETRIC LEAVES2 gene of *Arabidopsis thaliana* is required for formation of a symmetric flat leaf lamina, and encodes a member of a novel family of proteins characterized by cysteine repeats and a leucine zipper (Iwakawa *et al.*, 2002).

The ASYMMETRIC LEAVES2-like gene (ASL) family has genes distributed on several chromosomes. GenBank search revealed that *ASL37* (GenBank accession: AB080839), *ASL39* (GenBank accession: AB080841), *ASL40* (GenBank accession: AB080842) and *ASL41* are among these genes. *ASL41*, the gene downstream of the DsE insertion site in ET253, is located on chromosome 4, whereas *ASL37*, *ASL39* and *ASL40* are distributed on chromosomes 1, 5 and 3, respectively. Therefore, the Southern hybridization patterns shown in Figure 4.10 could represent these genes with the similarity to *ASL41* decreasing in the order of *ASL40*, *ASL39* and *ASL37* (data not shown). The mRNA of these genes can be prepared from *Arabidopsis* seedlings 12 days after sowing (Iwakawa *et al.*, 2002), indicating that these genes are mainly expressed during vegetative growth. Our Northern analysis (Figure 4.11) also confirmed that the *ASL* genes are widely expressed in different vegetative tissues. However, the slight difference of the size between mRNA detected in floral tissues and mRNA detected in leaf also suggests that different *ASL* genes may be expressed in different tissues. The sizes of mRNA from *ASL41*, *ASL40* and *ASL39* are 987, 1039, and 1121 bases, respectively. We speculate that the mRNA detected from the floral tissues could be the transcript of the *ASL41* gene but the mRNA detected from the floral tissues could be the transcript of the *ASL40* or *ASL39* gene.

4.3.6 The MITE-like repetitive sequences

Transposons, which were originally identified in maize (McClintock, 1950), have been found in all eukaryotes and prokaryotes. They are classified into two basic groups

based on their mode of transposition (Rio *et al.*, 1996). Class I transposons, also known as retrotransposable elements such as long terminal repeat (LTR) retrotransposons, long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs), are mobilized through an RNA intermediate (Eickbush, 1992). Class II transposons, such as Ac/Ds and Mutator-like elements (MULEs) are DNA transposons that move via a DNA form (Finnegan, 1989). As class I elements excise very rarely, they are thought to be more mutagenic than class II elements. However, only class I elements have been frequently found to be associated with normal eukaryotic genes (Wessler *et al.*, 1995). Although the roles of these elements in evolution remain to be clearly understood (Lonnig and Saedler, 1997), it is at least clear that they have the potential to serve as a major source of mutation (Surzycki and Belknap, 1999). In addition, they may also contribute specific regulatory sequences to genes (Wessler, 1996; Britten, 1997).

In *Arabidopsis,* transposable elements account for at least 10% of the genome, or about one-fifth of the intergenic DNA (The *Arabidopsis* Genome Initiative, 2000). The class I and class II mobile elements found in other plant genomes also well present in *Arabidopsis* but class I elements are less abundant compared to many plants with large genomes. In the past a few years, several families of short interspersed elements with terminal inverted repeats have been found in plants. These elements are grouped into different families which share many structural, but not sequence, similarities, and are named as miniature inverted-repeat transposable elements (MITEs). Their common features include small size, AT richness and T+A-based specific target sequence, no coding capacity but with the potential to form stable secondary structures and conserved terminal inverted repeats (Tu, 1997). Since MITEs share characteristics of both class I and class II elements, it is now still not known whether they move by a DNA intermediate or via a retrotransposition related mechanism and therefore they remain unclassified (Casacuberta *et al.*, 1998).

At present, several families of MITE-like repetitive sequences have been found in *Arabidopsis* genome (Casacuberta *et al.*, 1998; Surzycki and Belknap, 1999). The sizes of these repetitive sequences range from 300 bp to 600 bp and the copy numbers range from less than 10 to over 1000. The repetitive sequence located directly downstream of the DsE insertion site in ET253 (Figure 4.6) is shorter with the length of only 165 bp. The copy number of this sequence is around 40. However, this repetitive sequence has almost all the features related to MITEs as described in 4.2.3. It is therefore believed that this repetitive sequnce could be a new family of the MITE-like elements in *Arabidopsis*.

In conclusion, several candidates, including the two peroxidase genes *atpH21-1* and *atpH21-2*, the ASYMMETRIC LEAVES2-like protein 41 coding gene *ASL41* and the MITE-like element, have been obtained as possible host for harboring the megagemetophyte-specific enhancer(s). The DsE insertion site in ET253 is located in the probable intergenic region downstream of both the two proxidase genes and the *ASL41* gene. To actually identify the enhancer(s) and its association, the isolated DsE flanking sequence need to be tested *in vivo* to find out the responsible sequence(s) before the enhancer(s) are linked to particular gene(s) or transposable element(s).

CHAPTER 5

IDENTIFICATION OF THE MEGAGAMETOPHYTE-SPECIFIC ENHANCER (MGSE)

5.1 Introduction

In chapter 4, preliminary studies of the candidate genes and MITE-like element that may associate with the MGSE function were presented and discussed. The only way to unambiguously identify the MGSE element, however, is probably to fuse various parts of the DsE flanking sequences from ET253 to a reporter gene under the control of a minimal promoter and transform these constructs back into naïve (wild) *Arabidopsis* to test these sequences for MGSE function. Theoretically, this MGSE element could either be part of a tissue specific promoter or an independent enhancer. According to Maniatis *et al.* (1987), a promoter is required for accurate and efficient initiation of transcription, whereas an enhancer increases the rate of transcription. Promoters are located immediately upstream from the start site of transcription while enhancers can act on cis-linked promoters at great distances in an orientation-independent manner and can also function downstream from the transcription unit. Considering the fact that the DsE insertion site in ET253 is located in a region which is less likely to be the upstream region of a gene (see chapter 4), it is anticipated that the MGSE could also be an independent enhancer or controlling a remote gene. Larger DsE flanking region should therefore be searched and tested.

To test a DNA sequence for enhancer function, establishment of a reliable enhancer testing system is necessary. Since enhancer alone is not functional but can act on cis-linked promoters, the sequence to be tested for enhancer activity has to be fused with a minimal promoter (or called basal promoter or core promoter) and placed upstream of a reporter gene to form a testing unit (heterologous promoter). A minimal

promoter is the promoter sequence containing the TATA transcription initiation site. It is not functional alone but can be activated by other cis-elements or enhancers. The most commonly used minimal promoter for testing plant cis-element is the CaMV 35S minimal promoter (Benfey *et al.*, 1989; Benfey *et al.*, 1990a; Benfey *et al.*, 1990b). This minimal promoter has not only been used in enhancer trap systems (Sundaresan *et al.*, 1995; Klimyuk *et al.*, 1995; Campisi *et al.*, 1999), but also been successfully applied in testing cis-elements for temporal or spatial gene expression, such as an M phase-specific cis-acting element in plant B-type cyclin genes (Ito *et al.*, 1998) and pollen-specific elements in late anther tomato (LAT) genes (Eyal *et al.*, 1995).

The next step is to test the constructs *in vivo* to determine the function of certain DNA sequences. A high efficient and reliable plant transformation system is therefore needed for introducing constructed DNA into plants. The *Agrobacterium tumefaciens*-mediated gene transfer system is the most widely used system for plant transformation (Hooykaas and Schilperoort, 1992; Zupan *et al.*, 2000). Recently, the pCAMBIA binary vector series (www.cambia.org), developed from the pPZP family of *Agrobacterium* binary vectors (Hajdukiewicz *et al.*, 1994), has shown to be a powerful tool kit for the construction of T-DNA with appropriate constructs for transformation. For *Arabidopsis*, gene transfer used to be done by leaf disc transformation (Lloyd *et al.*, 1986) or root transformation (Valvekens *et al.*, 1988), which are all based on the use of *in vitro* cultured tissues. The newly invented *In planta* transformation method (Bechtold *et al.*, 1993) revolutionized the *Arabidopsis* transformation process by using bolted *Arabidopsis* plants as recipients. large number of stable transformants can easily be obtained after vacuum infiltration of the whole *Arabidopsis* plants with Agrobacterium inoculum. This method was further simplified as 'floral dip' by Clough and Bent (Clough and Bent, 1998a). To test and determine the MGSE sequence and function, big numbers of constructs need to be tested in *Arabidopsis*, this *In planta* transformation method is therefore used.

This chapter presents works related to the construction of transformation vectors for testing enhancer function, the identification of the MGSE element and the elucidation of some basic features of the MGSE element in *Arabidopsis*.

5.2 Results

5.2.1 The megagametophyte-specific enhancer is located at the upstream region of DsE insertion site in ET253

The strategy we used for the identification of the DsE flanking region for megagametophyte-specific enhancer activity was to divide the whole DsE flanking sequence obtained into several parts and test each part separately *in vivo* in wild type *Arabidopsis*. The two putative promoter regions of atpH21-1 and atpH21-2 genes, and the 2195 bp region between atpH21-1 and the DsE insertion site (Figure 4.6b) were chosen to test first.

Based on sequence alignment between the 5' regions of atpH21-1 and atpH21-2 genes, putative promoter sequences for these two genes were determined and named as Patp1 (1.28kb) and Patp2 (1.2kb), respectively. Primers (forward: ATP1-PT, 5'-GCCCAAGCTTGTTCCTTCTTCTTGTCTGCA-3' and ATP2-TP, 5'-GCCCAAGCTTGT-ATGAAGTTGTCCTCTTAGT-3'; reverse: ATP9a-PBm, 5'-GGTGCCATGGAGGAGAG-GTTGATGGCGAGA-3') were designed for specific amplification of these two putative promoter sequences plus part of the first exon (23 codons from the ATG start codon) for in-frame fusion with the *gusA* reporter gene. HindIII and NcoI sites were added in the forward and reverse primers respectively so that the PCR fragments can be cloned directly in the right orientation. PCR were carried out to amplify the two putative promoter sequences with the primer pair ATP1-PT/ ATP9a-PBm for the promoter of atpH21-1 gene and the primer pair ATP2-PT/ ATP9a-PBm for the promoter of atpH21-2 gene. The PCR products were digested with HindIII and NcoI, and the HindIII/NcoI

fragments were then cloned into promoter testing vector pCAMBIA1281Z (Genbank accession: AF234294), resulting in pWY-H80 and pWY-H84 (Figure 5.1).

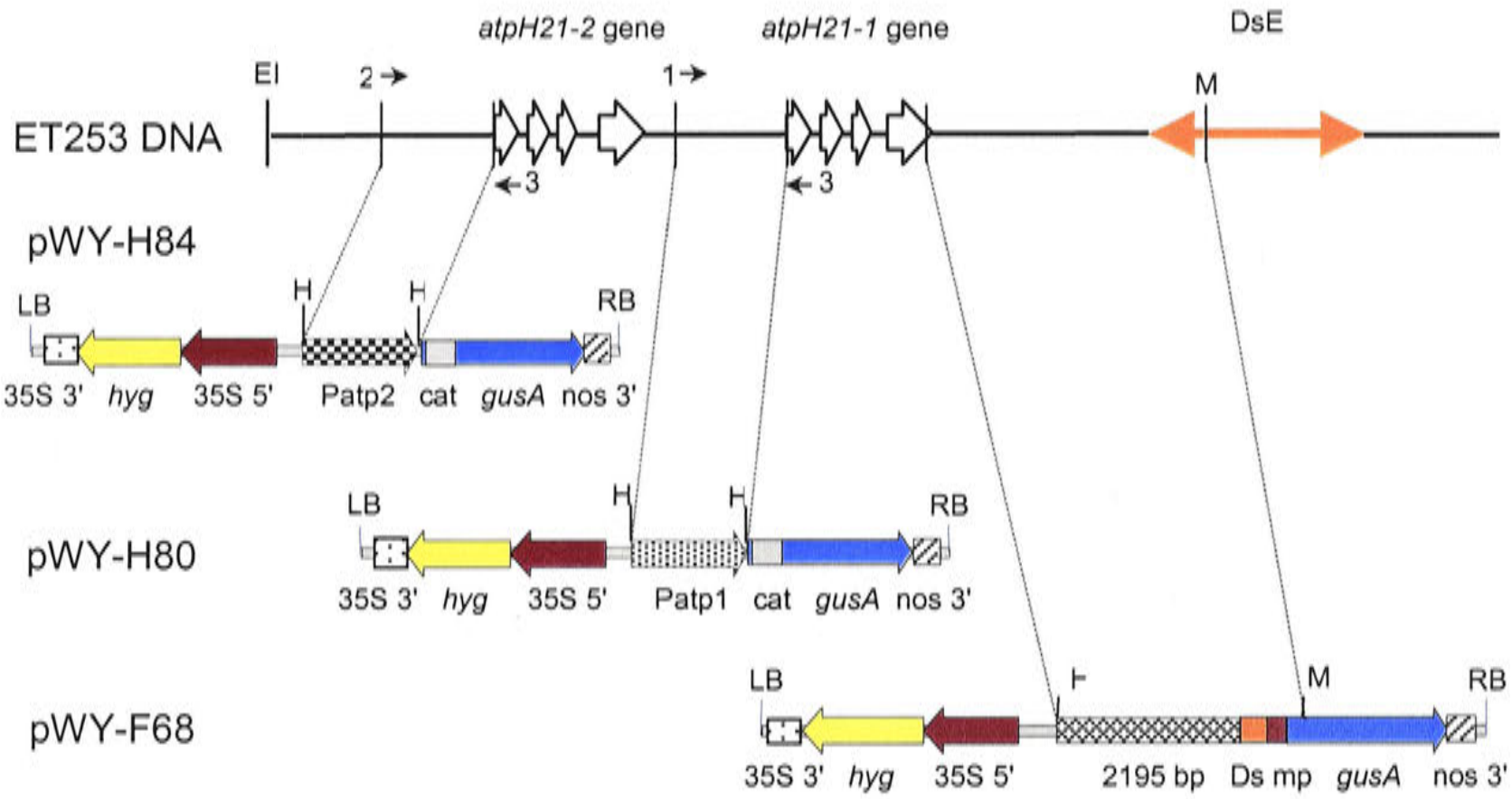


Figure 5.1 DsE flanking region in ET253 and the schematic maps of T-DNA region in pWY-H80, pWY-H84 and pWY-F68.

The DsE flanking region in ET253 is depicted as the solid line and the detailed structure of the DsE is referred to Figure 4.1a. The positions of the DNA fragments to be tested for MGSE activity in constructs pWY-H80, pWY-H84 and pWY-F68 are indicated by doted lines. Primers used for generating Patp1 and Patp2 are represented by small arrows (1, ATP1-PT; 2, ATP2-PT; 3, ATP9a-PBm). The annotations for other elements in the T-DNA are: *gusA*, β-glucuronidase gene coding region; cat, catalase gene intron; Ds, 3' Ds sequence (221bp) from the original enhancer trap DsE; mp, 35S minimal promoter (-1 to -46 region of the CaMV 35S prompter); nos3', nopaline synthase gene terminator; 35S 5', CaMV 35S promoter; 35S 3', CaMV 35S terminator; *hyg*, hygromycin resistance gene; LB, T-DNA left border; RB, T-DNA right border. The abbreviations for restriction enzymes are: EI, EcoRI; H, HindIII; M, MfeI.

For testing the 2195 bp direct upstream flanking sequence of the DsE insertion site, the 2.8kb fragment between the HindIII site at the end of the 4th exon of atpH21-1 gene and the MfeI site in *gusA* gene in DsE was placed in pCAMBIA1201 with the

same enzyme cutting sites. The derivative construct is pWY-F68 (Figure 5.1). The important point here is that the structure of the 2.8kb fragment is exactly the same as that in original ET253 line. This ensures the recurrence of the GUS expression pattern if the MGSE is located in this 2195 bp region.

Figure 5.1 shows the T-DNA regions of pWY-H80, pWY-H84 and pWY-F68. The location of Patp1, Patp2 and the 2195 bp fragment are illustrated. These constructs were transformed into wild type Landsberg erecta and different tissues from individual T_1 line was collected and subjected to GUS staining for determining any GUS expression.

Table 5.1 GUS expression in different tissues of pWY-F68 transformed lines of *Arabidopsis*

Tissues	pWY-F68 transformed T1 lines																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
embryo sac	++	++	-	+	+	+	-	-	+	+	+	+	++	-	-	+	++	++	+	+
carpel	+	-	-	+	-	-	-	+	+	-	+	+	+	-	-	+	-	-	-	-
stigma	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
style	+	-	-	+	-	+	-	+	+	+	-	+	+	-	-	+	-	-	+	+
anther	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	+	+	+	++	+
filament	+	-	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-
petal	+	-	-	+	+	+	-	+	+	-	+	-	+	-	-	+	-	-	-	-
sepal	+	-	-	+	-	-	-	+	+	-	+	-	+	-	-	+	-	-	-	-
leaf	+	-	+	+	+	+	-	+	++	+	+	+	+	-	-	+	-	-	-	-

+, GUS expression observed

++, very strong GUS expression observed

-, no obvious GUS expression observed

From eight pWY-H80 transformed T_1 lines, no GUS expression was observed in all tissues examined including floral tissues and leaf (root was not included) while six out of eleven pWY-H84 transformed T_1 lines showed GUS expression specifically in

anthers (mainly in pollen grains) (data not shown). However, from twenty pWY-F68 transformed T_1 lines, fifteen lines revealed GUS expression in the embryo sac (Table 5.1). Some lines, such as lines 1, 2, 13, 17 and 18, showed very strong GUS staining in the embryo sac. This clearly indicates that the 2195 bp direct DsE flanking sequence possesses the MGSE activity. An example of ovules of a pWY-F68 transformed T1 plant with GUS expression in the embryo sac is shown in Figure 5.2.

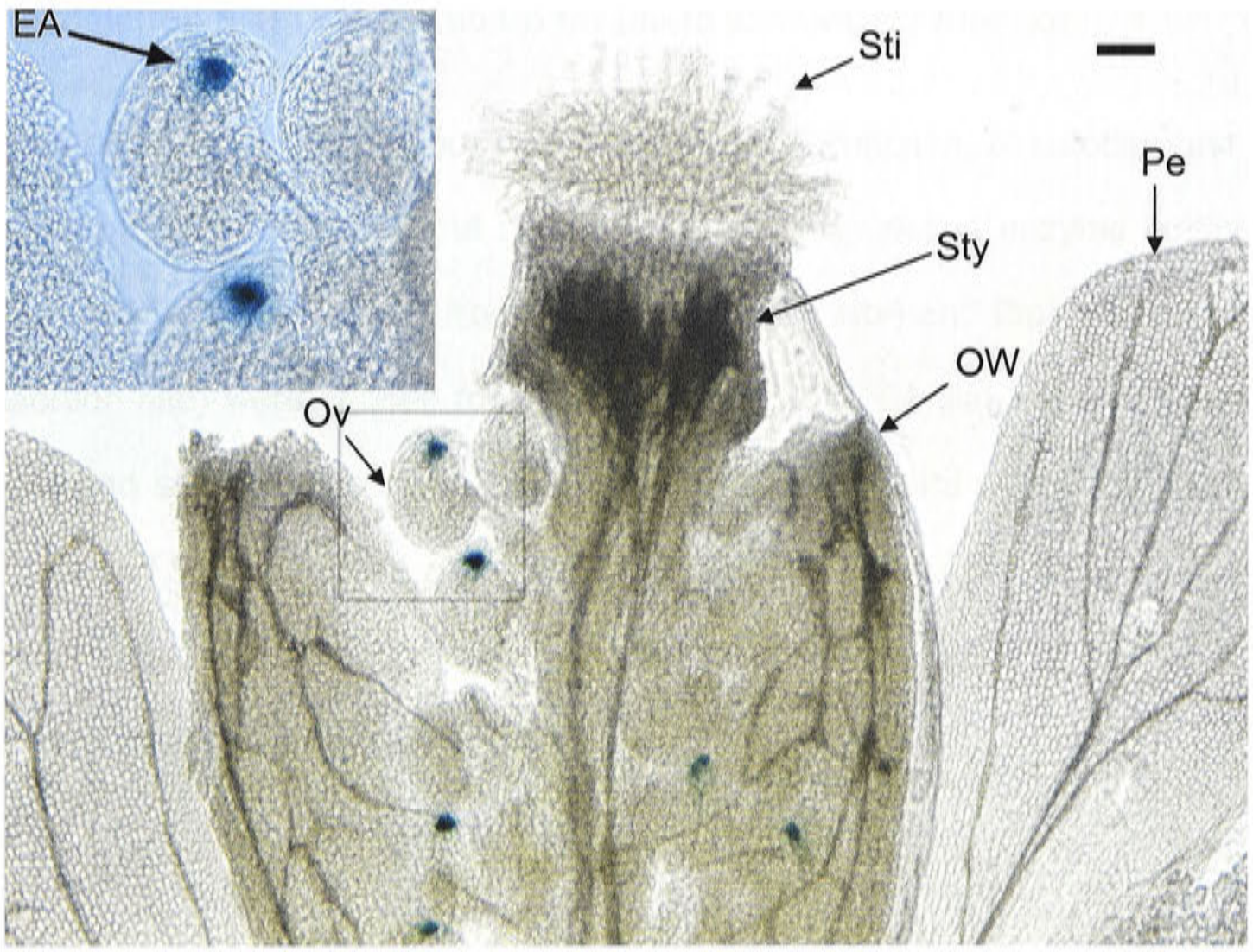


Figure 5.2 An example of pWY-F68 transformed T1 line showing GUS expression in megagametophyte.

The flowers from pWY-F68 transformed T1 plants were mounted onto a slide in clearing solution (20% lactic acid, 20% glycerol, 1×PBS) after GUS staining and clearing as described in 2.8. Ovules were exposed by placing a cover slip and pressing lightly to rapture the ovary wall. This picture shows a flower from line 2 with only strong GUS expression in the megagametophyte (mainly in the egg apparatus). Flowers with strong GUS expression in other tissues are shown in table 5.1. EA, egg apparatus; Ov, ovule; OW, ovary wall; Pe, petal; Sti, stigma; Sty, style. Bar = 50 μm.

Surprisingly, however, most of the pWY-F68 transformed lines also showed strong GUS staining in tissues other than the embryo sac. The GUS staining patterns of the 20 pWY-F68 transformed T_1 lines are listed in Table 5.1. The tissues with GUS expression are mainly leaf, style, carpel, petal, sepal and anther. The obvious background GUS expression pattern suggests that some cis elements are influencing the GUS gene expression.

5.2.2 5' deletion test of the 2195 bp fragment with MGSE function

To narrow down the region responsible for the MGSE function, 5' deletion test of the 2195 bp fragment was carried out by using the easy restriction enzyme cutting sites within the fragment. SpeI (1.3kb from the DsE insertion site) and BlpI (318 bp from the DsE insertion site) were chosen for this purpose. pWY-F68 was either digested with XbaI/SpeI and self ligated to generate pWY-J26.2, or digested with *Hind*III/BlpI, blunt ended, and self ligated to form pWY-J47.3 (Figure 5.3). These two deletion constructs were transformed into wild type Landsberg erecta to test the ability of the two deletion fragments in driving GUS expression.

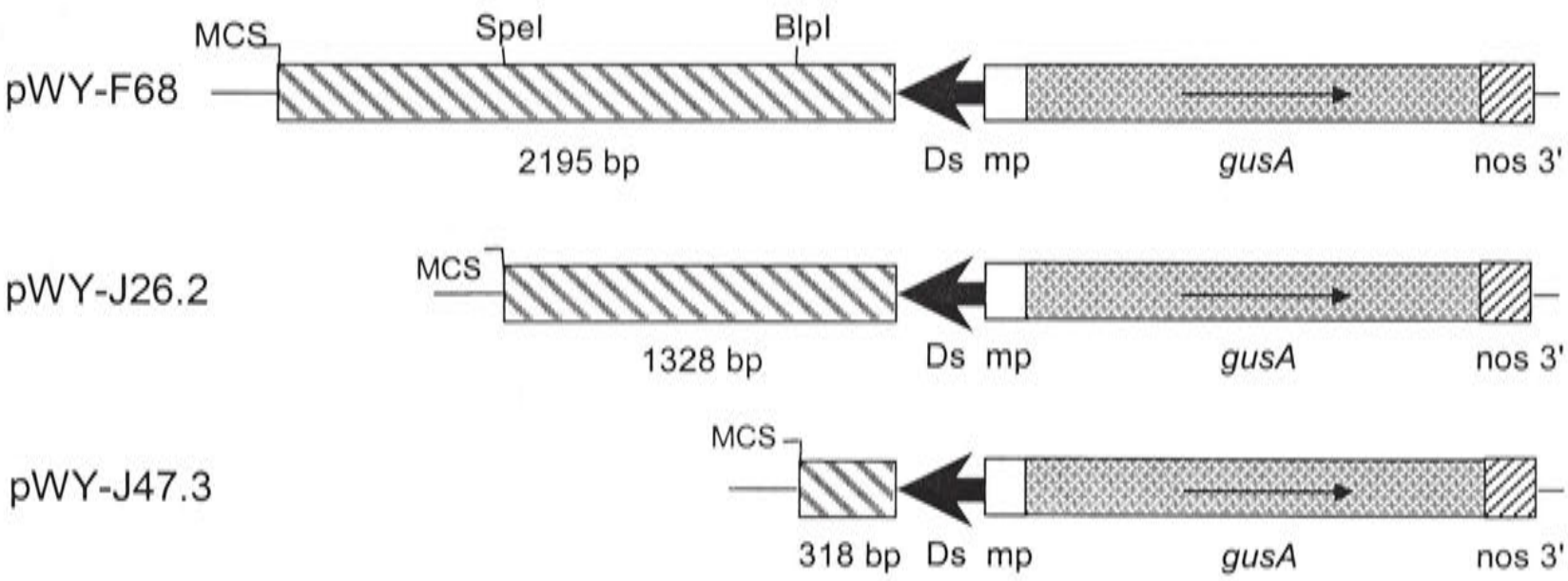


Figure 5.3 5' deletion test of the 2211bp upstream flanking sequence of the DsE insertion site in ET253.

5' deletions of the 2211bp sequence in pWY-F68 were carried out using existing unique SpeI and BlpI sites inside the sequence. pWY-J26.2 was constructed by double digestion of pWY-F68 with XbaI in the multiple cloning site (MCS) and SpeI, and religation; pWY-J47.3 was constructed by double digestion of pWY-F68 with HindIII in the MCS and BlpI, blunting both ends using T4 DNA polymerase and religation. The sizes of the resulting DsE flanking sequences in pWY-J26.2 and pWY-J47.3 are 1328bp and 318bp, respectively. Ds, 3' Ds sequence (221bp); mp, 35S minimal promoter (-1 to -46 region of the CaMV 35S prompter); nos3', terminator of nopaline synthase gene.

Table 5.2 GUS expression patterns of pWY-J26.2 and pWY-J47.3 transformed lines

Tissues	pWY-J26.2 transformed T1 lines													pWY-J47.3 transformed T1 lines						
	1	2	3	4	5	6	7	8	9	10	11	13	14	1	2	3	4	5	7	8
embryo sac	+	-	++	+	-	-	-	++	-	-	-	+	++	+	++	-	+	-	+	+
carpel	+	-	-	+	+	-	-	-	-	-	-	+	-	-	-	+	-	-	+	+
stigma	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
style	+	-	-	+	-	-	-	+	++	-	++	+	-	-	+	++	+	++	+	++
anther	-	-	++	-	-	-	-	+	-	-	+	-	-	-	+	++	+	++	+	+
filament	+	-	-	+	+	-	-	-	+	-	+	++	-	-	-	++	-	++	-	++
petal	+	-	-	+	+	-	-	-	-	-	-	+	-	-	-	+	-	+	+	++
sepal	+	-	+	+	-	-	-	-	+	-	+	++	-	+	+	++	+	+	+	+
leaf	++	-	+	+	-	-	-	+	+	-	+	+	-	+	+	+	-	+	+	+

+, GUS expression observed

++, very strong GUS expression observed

-, no obvious GUS expression observed

GUS staining results of different tissues from the T1 transformants of both pWY-J26.2 and pWY-J47.3 was summarized in Table 5.2. It can be seen from the table that the overall GUS expression patterns of the transformants from both deletion constructs are still quite the same as that from the pWY-F68 transformants (Table 5.1). Six of Fourteen pWY-J26.2 transformed lines and five of eight pWY-J47.3 transformed lines showed obvious GUS staining in the embryo sac, indicating that the 318 bp direct DsE flanking sequence is enough for conferring the GUS expression in embryo sac.

However, the background GUS staining pattern is more or less the same as pWY-F68 transformed plants.

5.2.3 Problems associated with the CaMV 35S promoter

The GUS staining results obtained from the transgenic lines of pWY-F68, pWY-J26.2 and pWY-J47.3 all revealed the same pattern: GUS is not only expressed in the embryo sac, but also in other parts of the plant (mainly sepal, petal, carpel, stamen, leaf, etc., root was not tested). The GUS expression in embryo sac should be due to the presence of the MGSE, but the consistent GUS expression in other parts of the plant could well be due to the effect from the CaMV 35S promoter controlling the *hyg* gene in the same T-DNA.

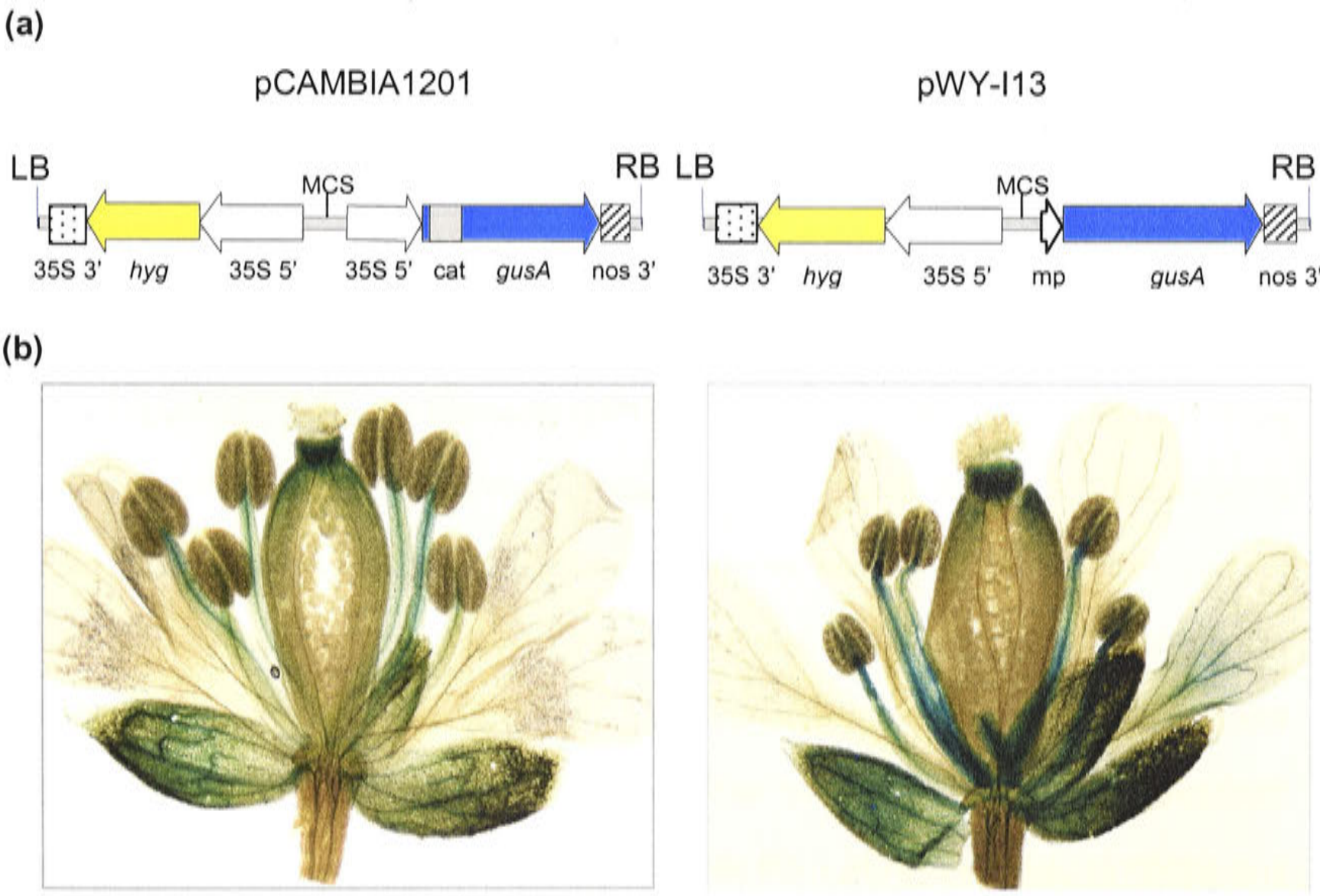


Figure 5.4 Similar GUS expression patterns from pCAMBIA 1201 and pWY-I13 transformed *Arabidopsis* plants.

(a) T-DNA region of pCAMBIA1201 and pWY-I13. *gusA*, β-glucuronidase gene; *hyg*, hygromycin resistance gene; 35S 5', CaMV 35S promoter; mp, CaMV 35S minimal promoter; 35S 3', CaMV 35S terminator; nos 3', terminator of nopaline synthase gene; cat, catalase gene intron; MCS, multiple cloning sites.

(b) GUS expression patterns of pCAMBIA1201 and pWY-I13 transformed T1 plants. Flowers from the pCAMBIA1201 and pWY-I13 transformed T1 plants were stained in GUS staining solution at 37°C for 18 hours and cleared in clearing solution for 2 hours before examination under microscope.

To confirm this speculation, plasmids pWY-I13 and pWY-K35.1 (the same as pWY-I13 except for the presence of the catalase intron in the *gusA* gene) were constructed to compare with pCAMBIA1201. The T-DNA of pWY-I13 or pWY-K35.1 has the *gusA* gene under the control of only the 35S minimal promoter and the full 35S promoter driving the hyg gene is placed in the opposite direction (Figure 5.4a).

After transforming wild type Landsberg erecta, hygromycin-resistant transformants were selected and examined by GUS staining. Result showed that the pWY-I13 and pCAMBIA1201 transformed plants demonstrated the same GUS expression pattern (Figure 5.4b). pWY-K35.1 transformed plants showed the same GUS expression pattern as well and therefore the result was not shown. This confirmed that the upstream promoter elements (UPEs) in the full 35S promoter could still show strong cis-effect on the 35S minimal promoters even though the full 35S promoter is placed in opposite orientation to the minimal promoter. Moreover, this cis-effect can not be blocked by spacing the full 35S promoter and the minimal promoter in certain distance (such as in pWY-F68, which has over 2kb space between the full 35S promoter and the minimal promoter). On the other hand, the observation of the GUS staining patterns of more than 50 transformants from pCAMBIA1201, pWY-I13 and pWY-K35.1 also clearly revealed that the UPEs from the 35S promoter do not show any activity in driving GUS expression in ovules. As a result, this also confirms that the

GUS expression in embryo sac is not influenced by the UPEs in the full 35S promoter but is controlled by other cis-element(s).

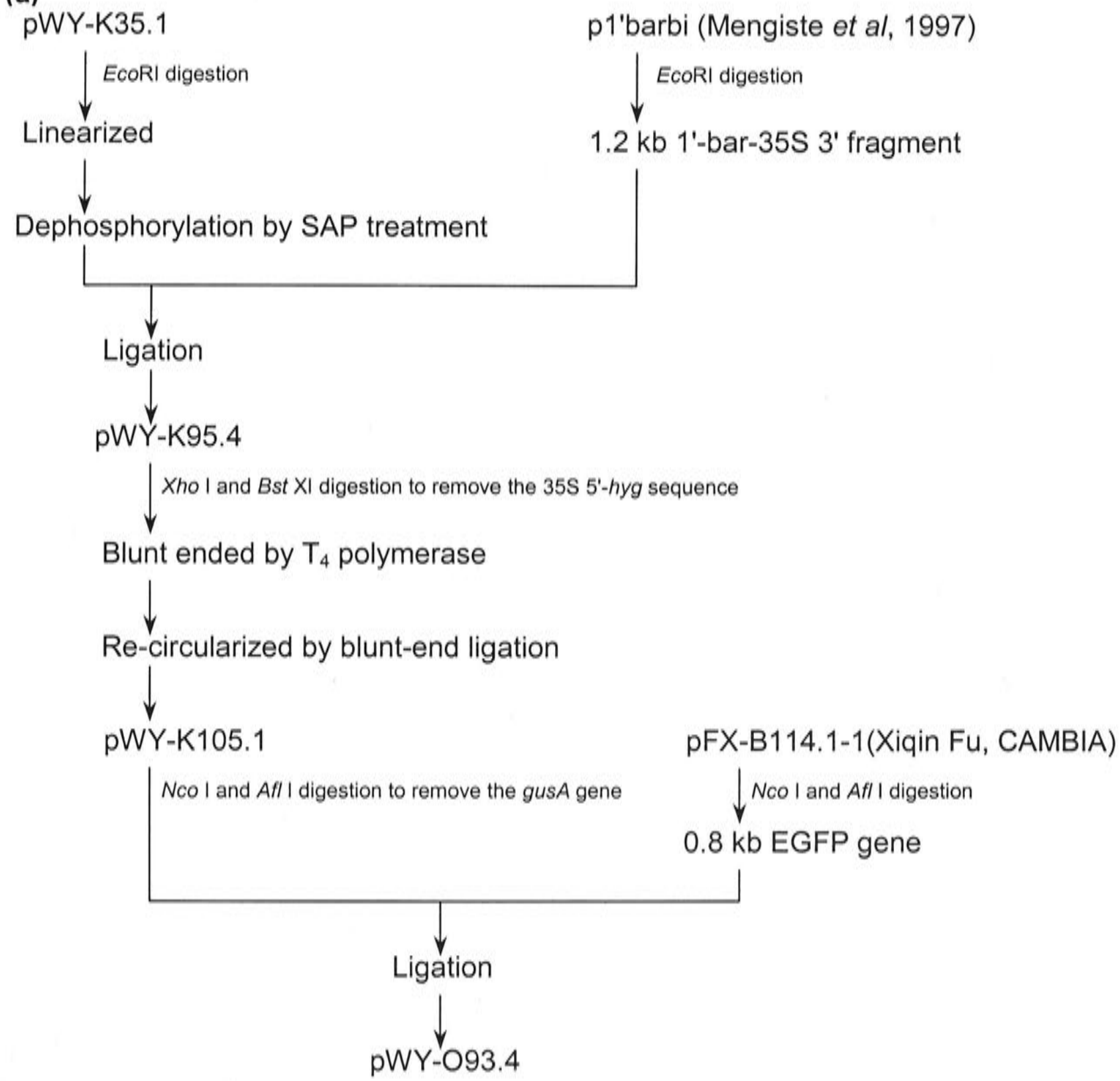
5.2.4 Construction of vectors suitable for testing enhancer function

After facing the problem stated in 5.2.3, it was realized that pCAMBIA1201 (and its derivatives such as pCAMBIA1281Z etc.) was not suitable for the examination of certain genomic DNA fragments with tissue-specific enhancer function. To test such enhancers, new transformation vectors containing no full CaMV 35S promoter must first be constructed. Figure 5.5a shows the procedures for the construction of two new vectors pWY-K105.1 and pWY-O93.4. As shown in Figure 5.5b, Vectors pWY-K105.1 and pWY-O93.4 were specifically designed for testing tissue specific cis-regulatory elements in *Arabidopsis*.

The use of the *bar* gene (White *et al.*, 1990; D'Halluin *et al.*, 1992) as a selectable marker has the advantage of selecting infiltrated *Arabidopsis* seeds directly on soil without seed surface sterilization, and moreover, transformants can be identified by simply spraying herbicide Basta on seedlings. In both vectors, the *gusA* gene or the enhanced GFP (EGFP) gene (Yang *et al.*, 1996) is driven by the CaMV 35S minimal promoter but they can not be expressed without the help of an enhancer element. Insertion of DNA fragment with certain enhancer activity into the multiple cloning sites (MCS) directly upstream of the minimal promoter can lead to GUS or GFP expression in accordance with the enhancer function. A very important point here is the use of the 1' promoter (Velten *et al.*, 1984) in driving the bar gene for selection of transgenic plants. There are two reasons for choosing this promoter. Firstly, the combination of 1' promoter with *bar* gene has been shown to be highly efficient in obtaining transformants in *Arabidopsis* (Mengiste *et al.*, 1997). Secondly, the 1' promoter has been used to control the *nptII* gene as plant selectable marker in the original enhancer

trap construct (Sundaresan *et al.*, 1995) and showed no obvious cis-activity on the 35S minimal promoter.

(a)



(b)

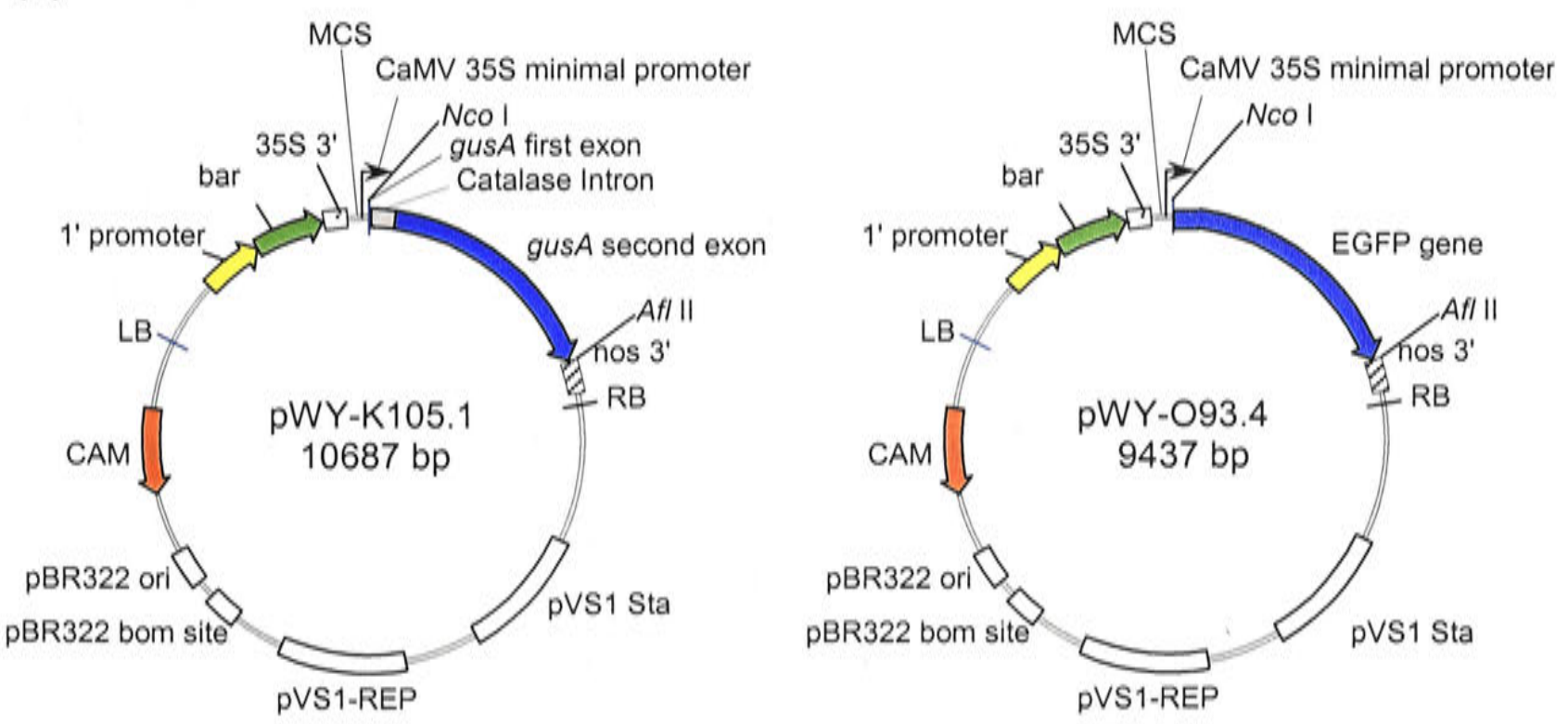


Figure 5.5 Construction of pWY-K105.1 and pWY-O93.4.

(a) Flow chart of the construction procedure for pWY-K105.1 and pWY-O93.4.

(b) Detailed maps of pWY-K105.1 and pWY-O93.4.

To test for any background GUS or EGFP expression in *Arabidopsis*, these two vectors were transformed into wild type *Arabidopsis*. Seeds were collected from the infiltrated plants and germinated on soil for Basta selection. After three rounds of Basta spray, 25 and 11 transformants were obtained from pWY-K105.1 and pWY-O93.4, respectively. The use of the 1' promoter in driving *bar* gene as selectable marker eliminates the interference of non-specific GUS expression from cis-activation by the strong CaMV 35S enhancer. This is confirmed by examining the GUS staining profile of the 25 pWY-K105.1 transformed plants and EGFP expression profile of the 20 pWY-O93.4 transformed plants. In contrast to the results from the pWY-I13 and pWY-K35.1 transformed plants that showed broad GUS expression, all pWY-K105.1 transformants had no detectable GUS expression in all tissues (floral, leaf and stem tissues) examined. The pWY-O93.4 transformed plants were all free of EGFP expression in all organs (tissues) except for the mature pollen grains, which showed some degree of auto-fluorescence (this is confirmed by comparing with wild type plants). Based on these results, the two new vectors appeared to be suitable for testing DNA fragment with specific enhancer activity. All the subsequent deletion tests and enhancer activity tests were then carried out based on the two vectors.

5.2.5 Further deletion test of the 318 bp fragment containing the MGSE

It was shown in 5.2.2 that the 318 bp sequence directly upstream of the DsE insertion site possesses the MGSE activity. A series of further 5' and 3' deletions were then designed to dissect the 318 bp sequence to identify the core sequence responsible for the MGSE enhancer activity (Figure 5.6). Deletion fragments were generated by PCR

amplification using the 1328 bp DsE flanking fragment in pWY-J26.2 (Figure 5.3) as template. For 5' deletions, primer ET-BL13-1 (the name is abbreviated as B1 in Figure 5.6a and the primer names given below follow the same way, 5'-GCCC**AAGCTT**GCCTTAATATCATACGAAAG-3') was paired with primers ET-TL13-1 (T1, 5'-GCCC**AAGCTT**GCTTAGCCTAATATCACAAA-3'), ET-TL13-2 (T2, 5'-GCCC**A-AGCTT**CGATATCATCTGAAACAGTC-3'), ET-TL13-3 (T3, 5'-GCCC**AAGCTT**CAAGA-TGTTTACGAACTTAC-3'), ET-TL13-4 (T4, 5'-GCCC**AAGCTT**GCAAACATTCGCG-ATTAACT-3'), ET-TL13-5 (T5, 5'-GCCC**AAGCTT**GAAGTAAAGTTACCCCAAGA-3') and ET-TL13-6 (T6, 5'-GCCC**AAGCTT**CCACGATGCAAATATATCGA-3') for the PCR amplification of a set of five deletion fragments with HindIII site (shown in bold letters) added to both ends for subsequent cloning. The derived deletion fragments were named as 5Δ318, 5Δ259, 5Δ206, 5Δ181, 5Δ136 and 5Δ77, respectively (Figure 5.6a). Four 3' deletion fragments 3Δ68, 3Δ117, 3Δ169 and 3Δ237 were generated in the same fashion using primer ET-TL13-1 to pair with primers ET-BL13-2 (B2, GCCC**AAG-CTT**GCATCGTGGAAATTTAAGTA-3'), ET-BL13-3 (B3, 5'-GCCC**AAGCTT**GGGGT-AACTTTACTTCA-3'), ET-BL13-4 (B4, 5'-GCCC**AAGCTT**GCGAATGTTTGCAATTA-GTA-3') and ET-BL13-5 (B5, 5'-GCCC**AAGCTT**GGGACTGTTTCAGATGATAT-3'), respectively (Figure 5.6a).

Specific PCR amplification of these 3' and 5' deletion fragments is shown in Figure 5.6b. The PCR fragments were then digested with HindIII and cloned into the HindIII site in the MCS upstream of the 35S minimal promoter in pWY-K105.1 (Figure 5.5b). Orientation of each cloned deletion fragment in pWYK105.1 was determined first by either NcoI/EcoRV digestion for the 3' deletions or NcoI/Psp1406I digestion for the 5' deletions, and finally by sequencing. The 5' deletion constructs were named as pWY-5Δ318 (positive control), pWY-5Δ259, pWY-5Δ206, pWY-5Δ181 and pWY-5Δ77, and the 3' deletion constructs were named as pWY-3Δ68, pWY-3Δ117, pWY-3Δ169, and pWY-3Δ237.

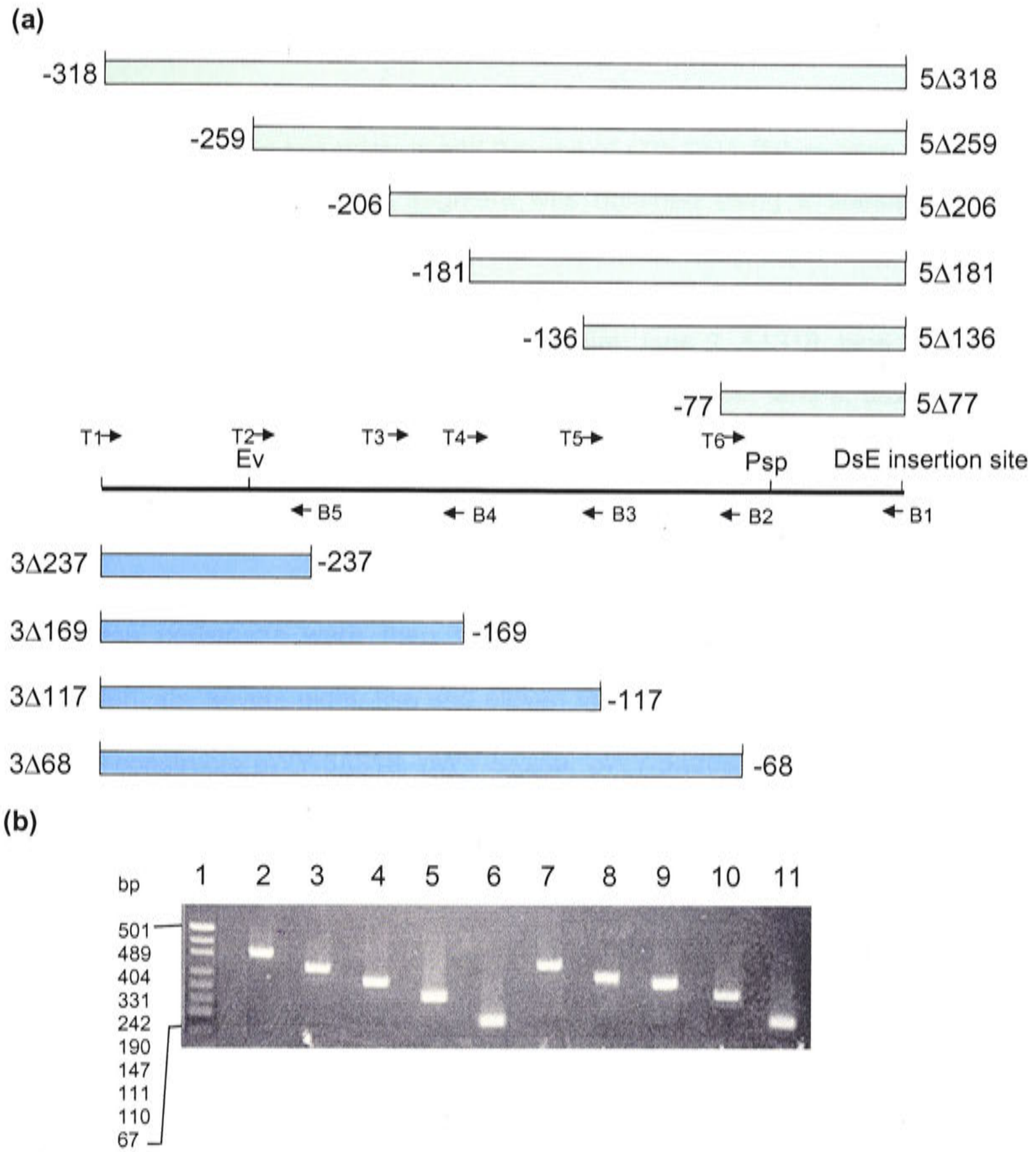


Figure 5.6 5' and 3' deletion of the 318bp DsE flanking sequence from ET253 for the determination of the MGSE activity.

(a) Strategy designed to generate a series of 5' and 3' deletion fragments from the 318 bp sequence. The line in the middle depicts the genomic region of the 318 bp sequence harboring the MGSE. Primer B1 was used to pair with T1 to T6 for PCR amplification of five 5' deletion fragments (boxes above the line) and the name of each derived fragment is shown on the right. Primer T1 was used to pair with B2 to B5 for the PCR amplification of four 3' deletion fragments (boxes below the line) and the name of each derived fragment is shown on the left. The numbers indicate the distance from the DsE insertion site. Ev, EcoRV; H, HindIII; Psp, Psp1406I.

(b) PCR amplification of the 5' and 3' deletion fragments. PCR was carried out using the GeneAmp XL PCR kit (see 2.3.3) in a 20 μl reaction mixture containing 10 ng of pWY-J26.2 DNA, 1×XL buffer II, 800 μM of dNTPs, 1.3 mM of $Mg(OAc)_2$, 0.5 μM of each of the two primers and 0.1 unit of *rTth* DNA Polymerase. Specific amplification of each deletion fragment was obtained using a standard PCR program (30 cycles of 30 sec denaturation at 94°C, 30 sec annealing at 55°C and 1 min extension at 72°C). PCR products were examined by electrophoresis on 3% agarose gel. The specific PCR products are: lane 2, 5Δ318; lane 3, 3Δ68; lane 4, 3Δ117; lane 5, 3Δ169; lane 6, 3Δ237; lane 7, 5Δ259; lane 8, 5Δ206; lane 9, 5Δ181; lane 10, 5Δ136; lane 11, 5Δ77. Lane 1 is the pUC19/HpaII DNA markers and their sizes are shown on the left.

These constructs were then tested in wild type *Arabidopsis*. After Basta selection, ten, six, seven, eight, five and eleven transformants were obtained from the 5' deletion constructs pWY-5Δ318, pWY-5Δ259, pWY-5Δ206, pWY-5Δ181 and pWY-5Δ77, respectively, and twelve, ten, nine and eleven transformants were obtained from 3' deletion constructs pWY-3Δ68, pWY-3Δ117, pWY-3Δ169, and pWY-3Δ237, respectively. GUS staining results showed that the majority of the transformants from all 5' deletion constructs have megagametophyte-specific GUS expression, but the transformants from all 3' deletion constructs revealed no GUS expression at all (Table 5.3). GUS expression in other floral tissues including sepal, petal, stamen, stigma and carpel, and other vegetative tissues such as leaf and stem were not detectable in all transformants. These results clearly suggest that the 77 bp sequence (5Δ77, Figure 5.6a) is sufficient to direct the megagametophyte-specific expression of the GUS gene. The deletion of 68 bp sequence from the 3' end of the 318 bp fragment abolished the MGSE function completely (3Δ68, Figure 5.6), indicating that the MGSE element is within the 77 bp region. The location of the 77 bp MGSE-containing sequence in the 11.5 kb DsE flanking region covered by BAC clone F19F18 (GenBank accession No: AL035605) on chromosome 4 is shown in Figure 5.7.

Table 5.3 Megagametophyte-specific activity of 5' and 3' deletions in driving GUS expression

Constructs	Transformed T1 line											
	1	2	3	4	5	6	7	8	9	10	11	12
pWY-5Δ318	-	M	M	-	M	-	M	M	-	M		
pWY-5Δ259	M	M	-	-	M	-						
pWY-5Δ206	-	M	-	M	M	M	-					
pWY-5Δ181	M	-	M	M	-	M	M	-				
pWY-5Δ136	M	M	-	M	-							
pWY-5Δ77	M	M	M	-	M	-	-	M	M	-	M	
pWY-3Δ237	-	-	-	-	-	-	-	-	-	-	-	
pWY-3Δ169	-	-	-	-	-	-	-	-	-			
pWY-3Δ117	-	-	-	-	-	-	-	-	-	-		
pWY-3Δ68	-	-	-	-	-	-	-	-	-	-	-	-
	GUS stained tissue											

M, megagametophyte; -, no GUS expression

* GUS stained tissue includes all floral tissues, stem and leaf. Root was not tested

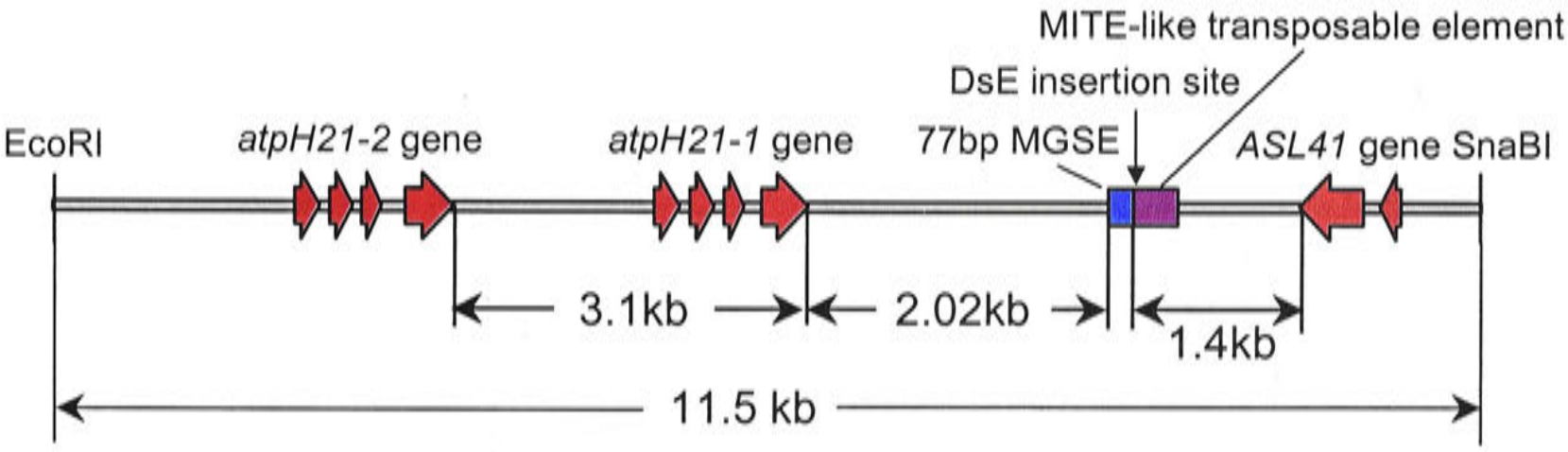


Figure 5.7 Location of the 77 bp MGSE-containing sequence in the 11.5 kb DsE flanking region covered by BAC clone F19F18 on chromosome 4.

The location of the 77 bp sequence covering the MGSE was determined by 5' and 3' deletion test. The approximate distances of this sequence to its nearby genes were shown in kilo base pairs.

In an attempt to further define the MGSE core sequence, the 77 bp sequence was scanned for MGSE activity by five 30 bp fragments plus one 27 bp fragment with 20 bp overlapping sequence between the adjacent fragments (Figure 5.8). This arrangement could define a sequence as small as 10 bp if the MGSE core sequence is within 30 bp. Six pairs of oligonucleotides were designed and synthesized for using as

either top strand or bottom strand to form double-stranded DNA fragment. These oligonucleotides are as follows: top strands, ET-TO66-1 (5'-**C**CACGATGCAAATATA-TCGATAACGTTATT**A**-3'); ET-TO66-2 (5'-**C**AATATATCGATAACGTTATTAAAAAAAG-T**A**-3'); ET-TO66-3 (5'-**C**TAACGTTATTAAAAAAAGTAACCGCATGAT**A**-3'); ET-TO66-4 (5'-**C**AAAAAAAGTAACCGCATGATATA-TTCTCTT**A**-3'); ET-TO66-5 (5'-**C**ACCGCAT-GATATATTCTCTTTCGTATGAT**A**-3'); ET-TO66-6 (5'-**C**ATATTCTCTTTCGTATGATA-TTAAGGC**A**-3'); bottom strands, ET-BO66-1 (5'-**AGCTT**AATAACGTTATCGATATATT-TGCATCGTG**GAGCT**-3); ET-BO66-2 (5'-**AGCTT**ACTTTTTTTAATAAACGTTATCGAT-ATATT**GAGCT**-3'); ET-BO66-3 (5'-**AGCTT**ATCATGCGGTTACTTTTTTTAATAACGTT-A**GAGCT**-3'); ET-BO66-4 (5'-**AGCTT**AAGAGAATATATCATGCGGTTACTTTTTTTT**GA-GCT**-3'); ET-BO66-5 (5'-**AGCTT**ATCATACGAAAGAGAATATATCATGCGGT**GAGCT**-3'); ET-BO66-6 (5'-**AGCTT**GCCTTAATATCATACGAAAGAGAATAT**GAGCT**-3'). *Sac*I and *Hind*III sites were created at 5' and 3' ends respectively to each fragment (*Sac*I to the 5' end and *Hind*III to the 3' end for the top strand, and the opposite for the bottom strand, as shown in bold letters) for directional cloning purpose. After annealing to form double-stranded DNA, each fragment was cloned into the MCS sites (between *Sac*I and *Hind*III sites) in pWY-K105.1 to generate constructs pWY-P5.1 to pWY-P5.6 (Figure 5.8).

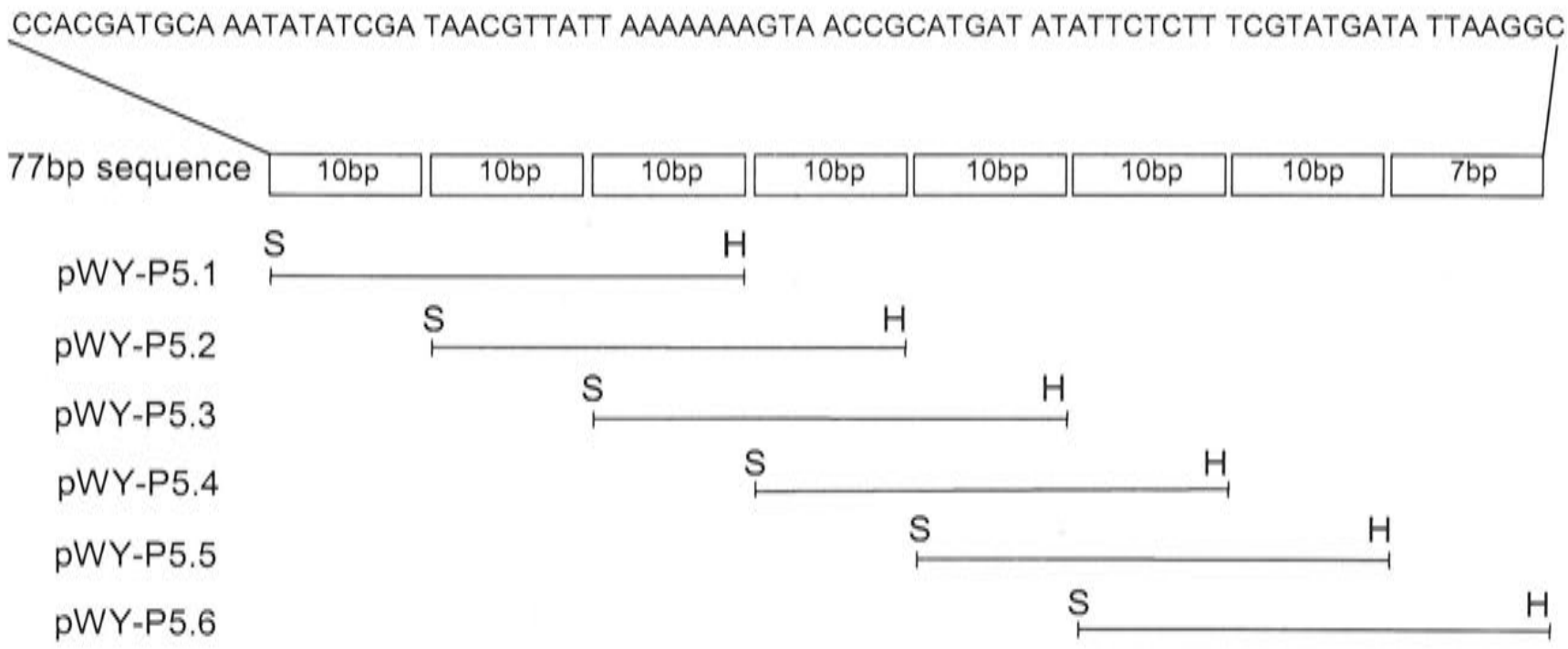


Figure 5.8 Further dissection of the 77bp sequence with MGSE activity.

The 77bp sequence with MGSE activity is presented on top. Six pairs of oligonucleotides, each covering 30 (27)bp, were designed and synthesized for the scanning of the 77bp sequence in six constructs. Sac I and Hind III sites were added to the 5' and 3' ends, respectively, in each pair for the directional cloning into the MCS sites in pWY-K105.1. The resulting constructs are pWY-P5.1 to pWY-P5.6 and the corresponding region covered by each construct is defined by the lines. H, Hind III; S, Sac I.

These constructs were then tested in wild type *Arabidopsis*. After Basta selection of the T1 seedlings, ten, thirty-one, eighteen, twenty-four, twenty-five and eighteen transgenic lines were obtained from pWY-P5.1, pWY-P5.2, pWY-P5.3, pWY-P5.4, pWY-P5.5 and pWY-P5.6, respectively. Unexpectedly, GUS staining of the above-ground tissues revealed that no GUS expression was detectable from all these transgenic lines. This result suggests that the MGSE core sequence is located in a region larger than 30 bp within the 77 bp fragment. Two possibilities exist here: either the protein factor binding site is bigger than 30 bp or the protein binding site consists of discrete DNA domains in the 77 bp sequence that work in a synergistic manner. Therefore, more work need to be done to further define the accurate domain(s).

5.2.6 The MGSE is unique in *Arabidopsis thaliana*

Since the MGSE was only tagged by the enhancer trap in just one locus, the possibility that the MGSE is present in *Arabidopsis* genome with multiple copies exists. If the MGSE is present in more than one locus, additional information about the possible role of the MGSE in the genome may be acquired by analyzing the MGSE at other loci. We therefore carried out BLAST search using the 77 bp MGSE sequence against the publicly available complete genomic sequence of ecotype Columbia but obtained no hits of the MGSE homologue at other locations.

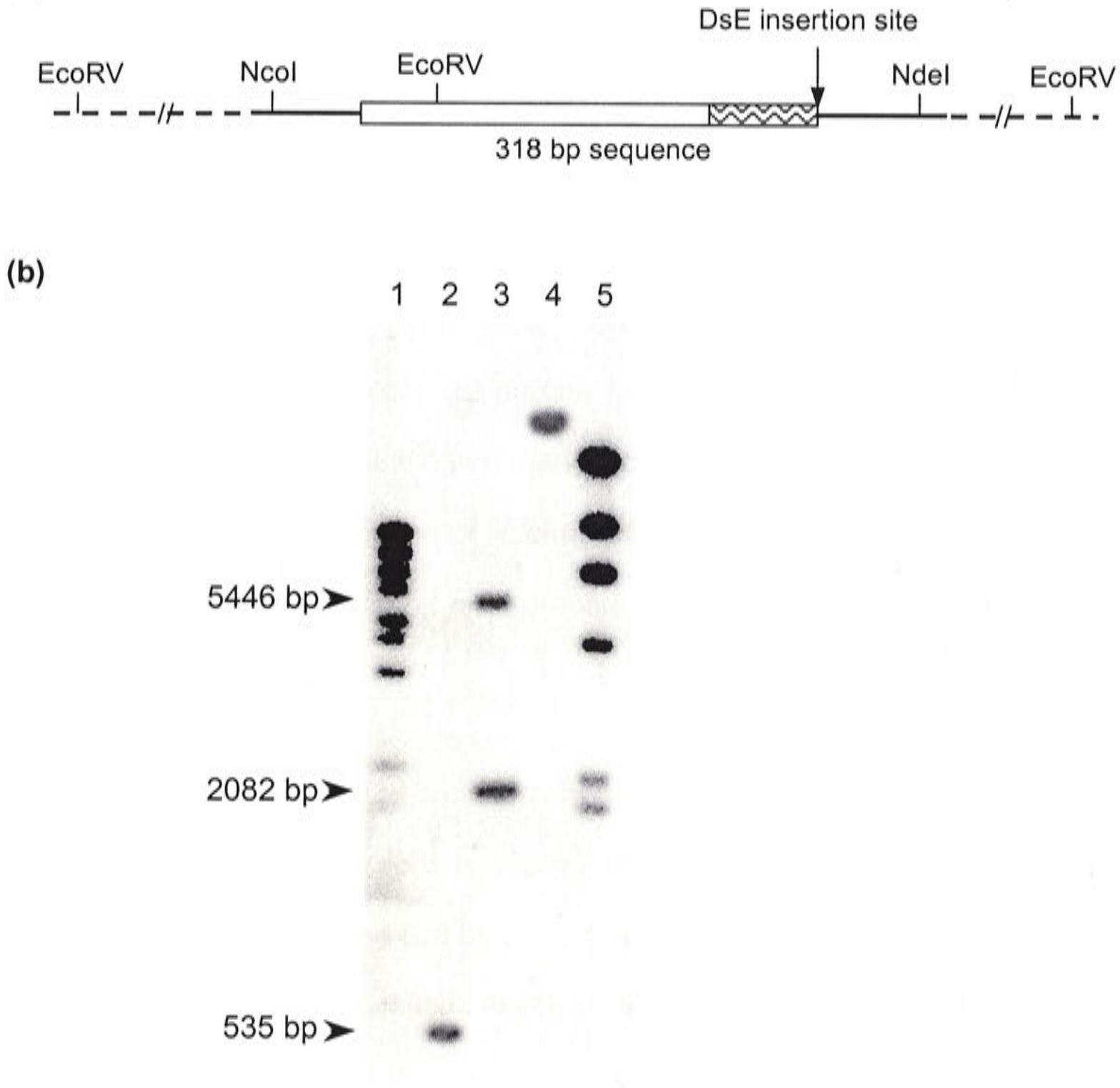


Figure 5.9 Single copy of the MGSE in *Arabidopsis* genome revealed by Southern blotting.

(a) Restriction map of the region surrounding the MGSE (not drawn in scale). Depicted are: chromosome DNA (lines, the doted lines represent longer distance); the 318 bp fragment used as probe in Southern hybridization (box, the zigzag part represents the MGSE sequence). The DsE insertion site is indicated by arrow.

(b) Southern blot of DNA from ecotype Landsberg erecta. Approximately 3 μg of genomic DNA from Ler was digested with NcoI/NdeI and EcoRV and probed with the 318 bp fragment. Lane 1, λ/BstEII DNA markers; lane 2, Ler DNA/NcoI+NdeI; lane 3, Ler DNA/EcoRV; lane 4, Ler DNA/undigested, lane 5, λ/HindIII. The sizes of the hybridized signals are indicated on the left side.

To make sure that it is also true for ecotype Landsberg erecta, Southern analysis was then used to determine the copy number of the MGSE using genomic DNA from Landsberg erecta. Since the 77 bp sequence is highly AT rich (68.8%), use the 77 bp sequence as probe may not give strong enough signal, the longer version, the 318 bp sequence (5.2.2 and 5.2.5), was then chosen as a probe for Southern hybridization. As shown in Figure 5.9a, there is one EcoRV site inside the 318 bp fragment, while NcoI and NdeI cut outside the fragment. These restriction enzymes were therefore chosen for cleavage of genomic DNA. Genomic DNA was digested with either *Eco*RV or the combination of *Nco*I and *Nde*I, separated on 1% agarose gel and blotted onto nylon membrane, and probed with the α-^{32}P-dATP labeled 318 bp fragment.

Figure 5.9b shows the Southern blotting result. When the DNA was cleaved with NcoI/NheI, the 318 bp probe hybridized only to a 535 bp fragment. When cleaved with EcoRV, the blot revealed two bands: 5446 bp and 2082 bp. These bands are all in right sizes according to calculations based on the detailed sequence information (4.2.3, chapter 4; Appendix B) and therefore represent only the region tagged by DsE. It is therefore confirmed that the MGSE has single copy in the *Arabidopsis thaliana* genome.

5.2.7 The MGSE is orientation independent

Enhancers are believed to be able to work in both orientations (Maniatis *et al.*, 1987). In 5.2.5, it was demonstrated that the two constructs pWY-5Δ318 and pWY-5Δ77 can confer megagametophyte-specific GUS expression. The next question is whether the MGSE can work in reverse orientation. To answer this question, two parallel constructs were made in which the 77 bp and the 318 bp fragments were replaced by their complementary copies (Figure 5.10).

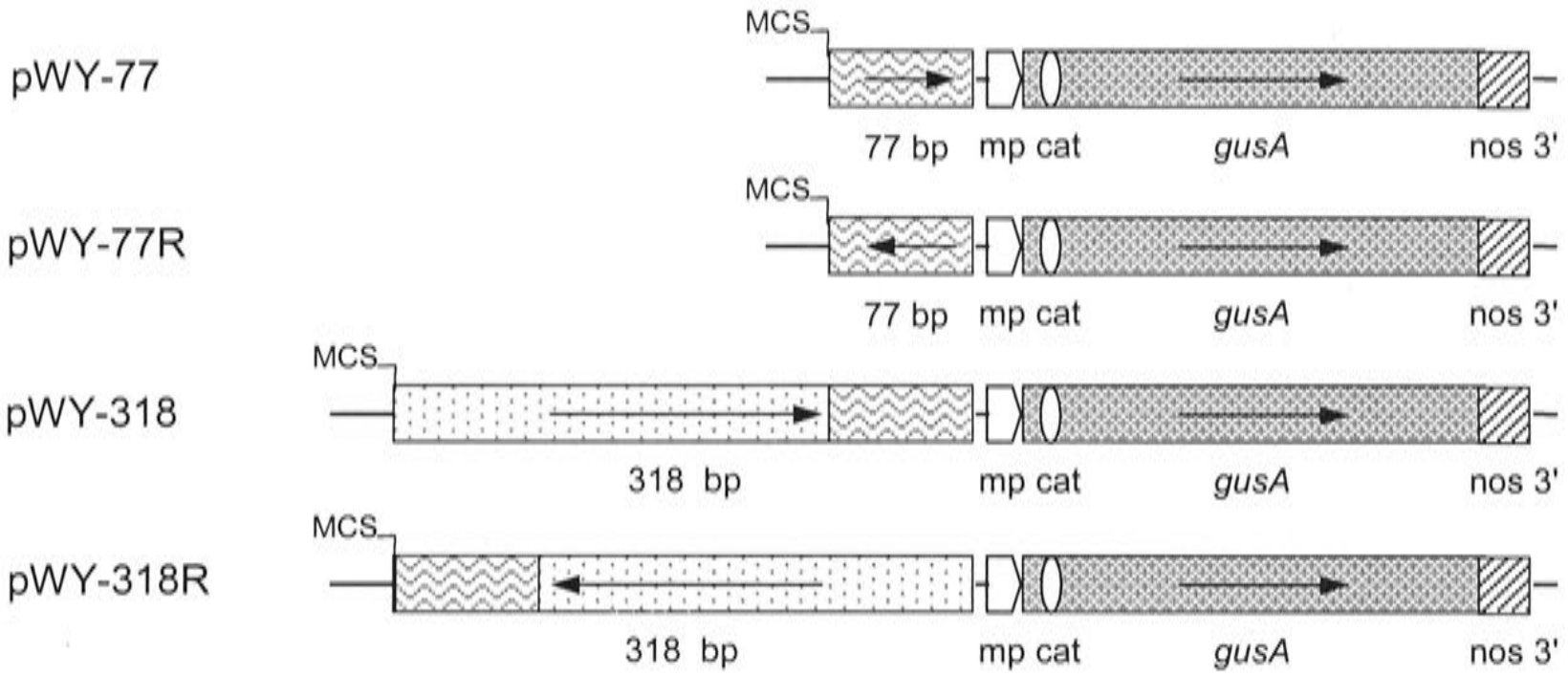


Figure 5.10 Constructs designed for testing the MGSE in two orientations.

The four constructs pWY-77, pWY-77R, pWY-318 and pWY-318R were constructed in vector pWY-K105.1 (5.2.4). Abbreviations are: mp, minimal CaMV 35S promoter; nos3', terminator of nopaline synthase gene; *gusA*, β-glucuronidase gene coding region; cat, catalase intron. 77 bp, the 77 bp sequence with MGSE activity; 318 bp, the 318 bp sequence with MGSE activity. The orientation of the 77 bp and the 318 bp sequence is indicated by arrows.

The orientations of the 77 bp sequence in pWY-5Δ77 and the 318 bp sequence in pWY-5Δ318 are the same as that from the original enhancer trap line ET253. To make the comparison easier, the names pWY-5Δ77 and pWY-5Δ318 were simplified here as pWY-77 and pWY-318 and their counterparts with the reversed 77 bp and 318 bp sequences were named as pWY-77R and pWY-318R respectively (Figure 5.10). As shown in Figure 5.11, seven of eleven pWY-5Δ77 transformed lines and eleven of sixteen pWY-5Δ77R transformed lines revealed megagametophyte-specific GUS expression, indicating that the MGSE can work in both directions. The visual comparison of the GUS staining in the megagametophytes of the pWY-5Δ77 and pWY-5Δ77R transformed plants also showed that the intensity of GUS staining in the pWY-5Δ77R transformed plants were generally higher than that in pWY-5Δ77 transformed plants in terms of both the level of GUS staining in each individual ovule and the

number of ovules showing GUS staining (data not shown). This suggests that the reverse copy of the 77bp sequence may have higher activity. In contrast, although six of ten pWY-5Δ318 transformed lines also revealed megagametophyte-specific GUS expression, none of the pWY-5Δ318R transformed lines (twenty-six lines examined) showed any GUS expression (Figure 5.11). It could be due to the existence of an insulator sequence upstream of the 77bp MGSE in the 318 bp fragment. This speculation needs to be further examined. However, the orientation dependent feature of the 318 bp fragment could be very useful when the MGSE is used to control certain gene expression.

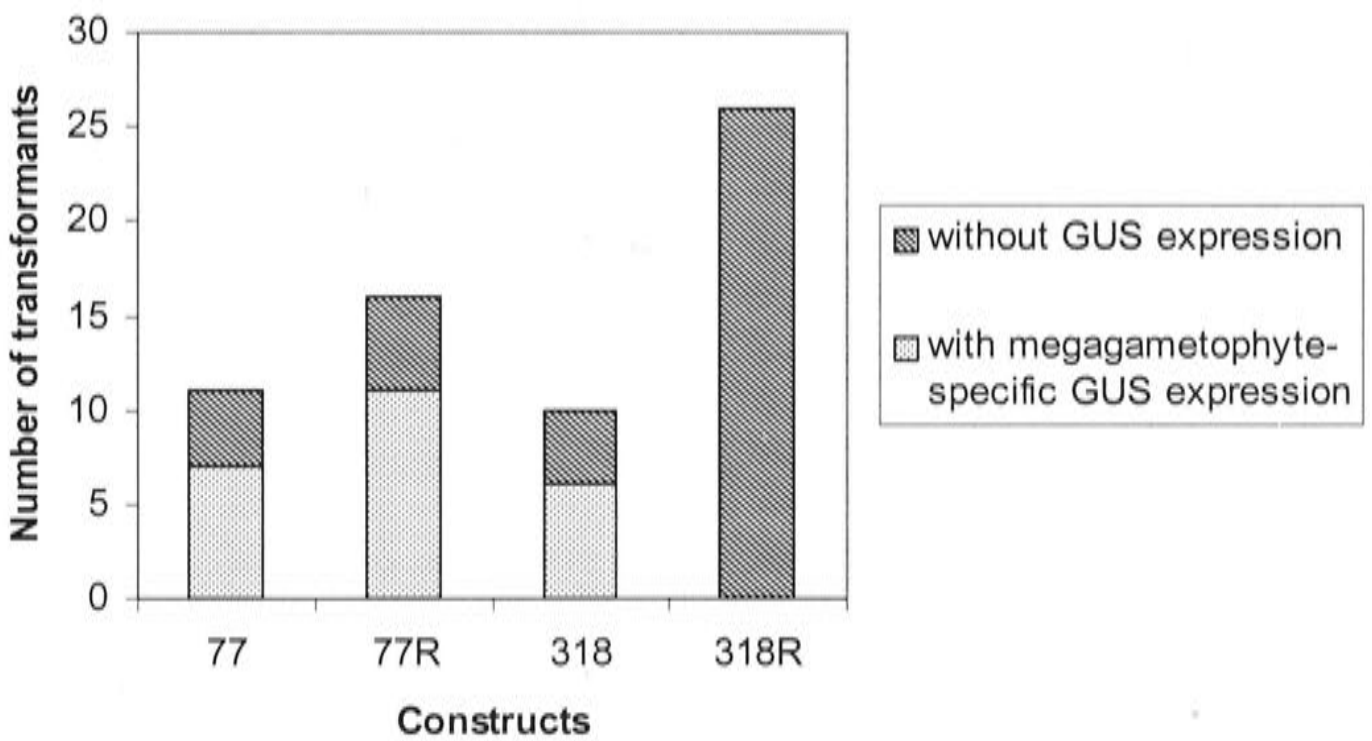


Figure 5.11 Test of the reverse copies of the 77 bp and the 318 bp fragments for megagametophyte-specific activity.

Transformants from the four constructs pWY-77, pWY-77R, pWY-318 and pWY-318R (abbreviated as 77, 77R, 318 and 318R, respectively) were examined by GUS staining of the floral tissues. Lines with specific GUS expression in embryo sacs were counted as GUS positive and lines without any detectable GUS staining were counted as GUS negative.

5.2.8 Tandem repeats of the MGSE can enhance its activity

The activity of cis-acting element sometimes can be increased by multiplying the element (Ito *et al.*, 1998). To test whether it is also the case for MGSE activity in relation to the possible application of the MGSE in driving certain gene expression at different levels by using different numbers of the MGSE, tandem repeats of the MGSE were tried and used together with a 35S minimal promoter to form an artificial promoter for driving GUS gene expression. To make such a construct, the 77 bp MGSE-bearing fragment was cut out from pWY-5Δ77 by HindIII digestion and inserted into the HindIII site in the MCS in pWY-K105.1 (Figure 5.5b). After transforming *E. coli* strain DH5α, colonies were randomly selected and plasmids were prepared and identified by NcoI/Psp1406I digestion. The colonies harboring plasmids with tandem repeats of the 77 bp fragment were selected against the colonies harboring plasmids with single insertion by NcoI/Psp1406I digestion patterns (plasmids with direct inverted repeats of the 77 bp sequence are lethal and therefore are eliminated naturally after transformation). The numbers of repeats of the 77 bp sequence were determined by sequencing. A plasmid with four tandem repeats of the 77 bp sequence in the same orientation as in pWY-5Δ77 was selected and named as pWY-5Δ77SS (Figure 5.12). In this construct, the *gusA* gene is under the control of the 35S minimal promoter and 4×77 bp MGSE fragment.

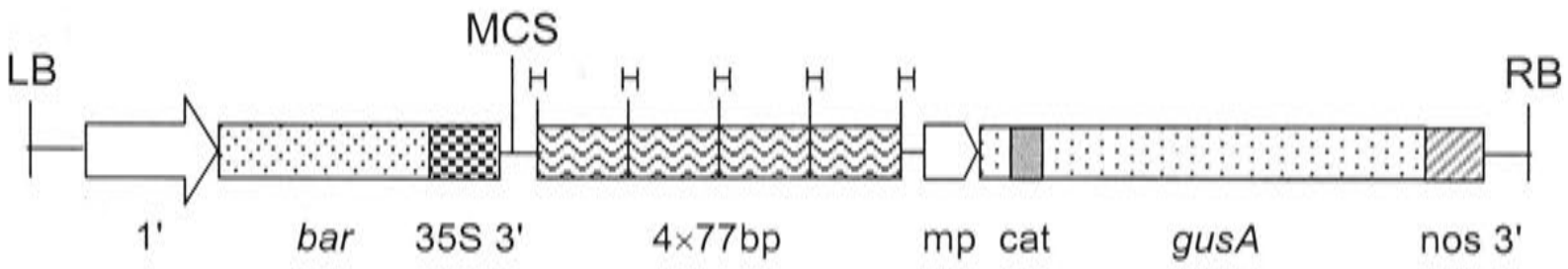


Figure 5.12 Schematic map of the T-DNA region of pWY-5Δ77SS.

The *gusA* gene is under the control of the artificial promoter consisting of the 35S minimal promoter (mp) and 4× tandem repeats of the 77bp MGSE sequence (4×77bp) in pWY-5Δ77SS. 1', 1' promoter; 35S 3', CaMV 35S gene terminator;

bar, bialaphos resistance gene; cat, catalase gene intron; nos 3', terminator of nopaline synthase gene; MCS, multiple cloning sites; H, Hind III;

pWY-5Δ77SS was also tested in wild type Landsberg erecta. Among the eight transgenic lines obtained, GUS staining result showed that six lines displayed stronger GUS expression in the embryo sac and the other two lines were with normal GUS expression level, which is mainly reflected by the numbers of ovules stained when compared with pWY-5Δ77 transformed plants. Unfortunately, no quantitative measurement can be done at the moment due to less tractable tissues in the ovules. However, it still suggests that there may be a dosage effect for the MGSE activity when different numbers of the element are present.

This dosage effect is also obvious when transgenic lines with different copy numbers of T-DNA insertions were compared by their GUS expression levels in the megagametophyte. To examine whether such dosage effects exist, genomic DNA was prepared from pWY-5Δ77R transformed T1 plants showing different levels of GUS expression and subjected to Southern blotting analysis after digestion with HindIII. Copy numbers of T-DNA insertion in the genome of each selected line was determined by using a 0.9 kb MfeI/BstBI fragment from the *gusA* gene, which contains no HindIII site, as probe. The GUS expression levels of these lines were then compared with their T-DNA copy numbers. The results are presented in Table 5.4. It can be seen from this table that the pWY-5Δ77R transformed lines with more T-DNA insertions tend to show higher levels of GUS expression. For example, the lines with 3-4 copies of T-DNA (lines 8, 10 and 11) obviously showed stronger GUS staining in the megagametophytes than the lines with single copy of T-DNA (lines 5 and 15). In the case of line 5, which showed no observable GUS expression, GUS could still be expressed but the level could be too low to be detected by GUS staining as described in Chapter 6. The results

presented in this table were only obtained from a small number of samples. To confirm this speculation, more samples need to be examined.

Table 5.4 The intensity of GUS expression and the T-DNA copy number(s) of some randomly selected pWY-5Δ77R transformed T1 lines

Line no.	No. of T-DNA insertion [a]	GUS expression in MG [b]
5	1	-
6	2	+
8	3	++
9	2	++
10	3	+++
11	4	+++
15	1	+

[a]The copy number of T-DNA insertion in each line was determined by Southern analysis using *gusA* gene as probe.

[b]The level of GUS expression in the megagametopytes (MG) of each line was evaluated by the combination of the intensity of GUS staining in each individual ovule and the number of ovules with visible GUS expression.

-, no visible GUS staining observed; +, weak GUS staining in only a few ovules within one gynoecium; ++, GUS staining in many ovules within one gynoecium; +++, strong GUS staining in most ovules within one gynoecium.

5.2.9 Megagametophyte-specific EGFP expression under the control of MGSE

In parallel, the capability of the MGSE in activating green fluorescence protein (GFP) gene expression was also examined. One advantage of the GFP gene over the GUS gene is that the expression of GFP is cell autonomous and can be detected in living cells which requires no exogenous substrates (Sheen *et al.*, 1995). This feature may be beneficial to better localize the reporter gene expression under the control of the MGSE.

In this experiment, the vector pWY-O93.4 (refer to 5.2.4) was used for the construction of construct with GFP gene driven by the MGSE. The GFP gene in pWY-O93.4 is a codon optimized GFP gene with dramatically enhanced GFP sensitivity and therefore named as EGFP gene (Yang *et al.*, 1996). Considering that the use of a single MGSE for driving the EGFP gene may have low level of EGFP expression, which could still be difficult to observe, the four tandem repeats of the 77 bp MGSE sequence (4×77 bp squence, refer to 5.2.8) was chosen to control the EGFP gene. The 4×77 bp MGSE fragment was inserted in the HindIII site within the MCS, which is upstream of the 35S minimal promoter in pWY-O93.4, to generate pWY-O93.1 (Figure 5.13a).

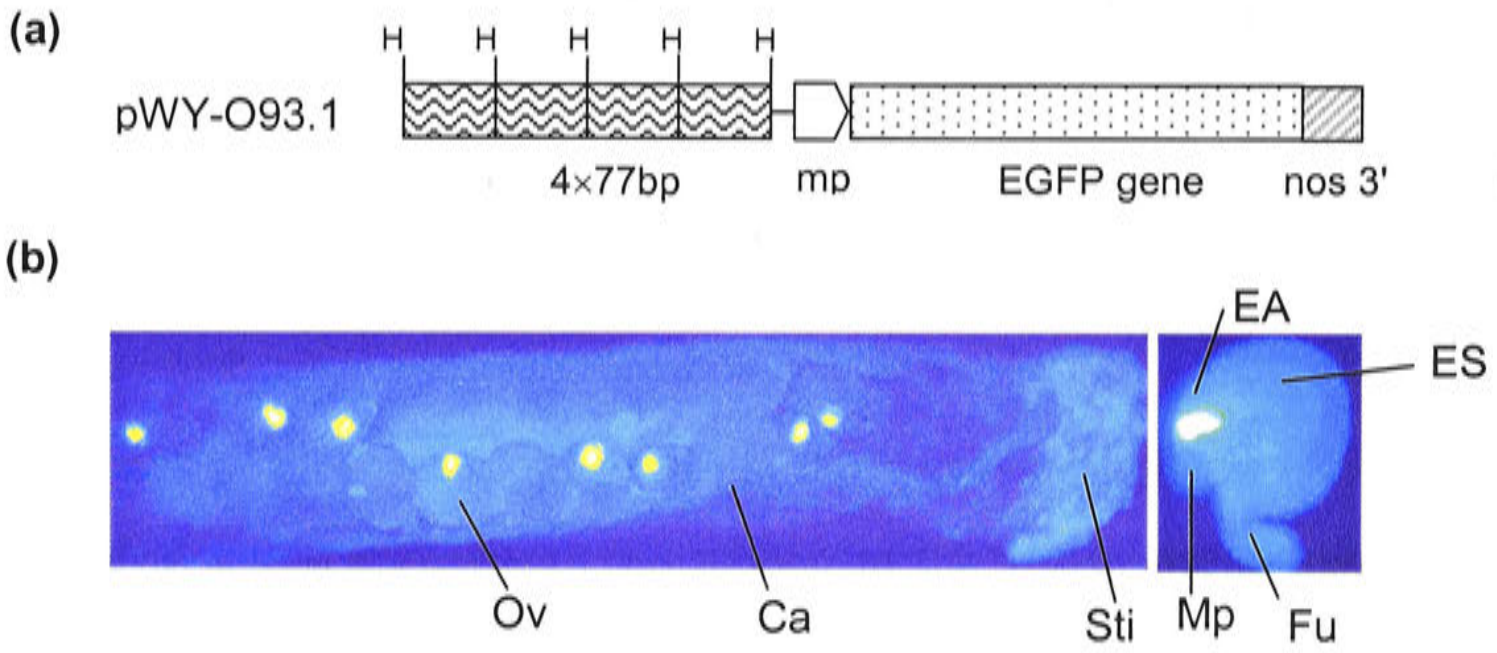


Figure 5.13 Specific GFP gene expression under the control of the MGSE.

(a) Schematic map of the EGFP expression construct in pWY-O93.1. The EGFP gene is under the control of the 35S minimal promoter plus 4× tandem repeats of the 77bp MGSE sequence. H, Hind III; mp, minimal promoter.

(b) Specific EGFP expression in the embryo sac of the pWY-O93.1 transformed *Arabidopsis* plant. The picture on the left shows a gynoecium with ovules expressing EGFP, and the picture on the right displays the location of the expressed EGFP in the embryo sac. Ca, carpel; EA, egg apparatus; ES, embryo sac; Fu, funiculus; Mp, micropyle; Ov, ovule; Sti, stigma.

As indicated in 5.2.4, result showed that the pWY-O93.4 transformed plants did not reveal any GFP expression in embryo sac (and other floral tissues as well). Although it was often observed that mature anthers display quite strong auto-fluorescence even in wild type *Arabidopsis*, other parts of the *Arabidopsis* plant are quite clean in auto-fluorescence (background). In contrast, in seven pWY-O93.1 transformed plants, five showed very strong and specific EGFP expression in embryo sac (mainly in egg apparatus). A typical EGFP expression pattern of pWY-O93.1 transformed plants is shown in Figure 5.13b. This EGFP expression pattern is consistent with the GUS expression pattern in pWY-5Δ77 and pWY-5Δ77SS transformed plants.

5.2.10 The existence of the MGSE in different ecotypes of *Arabidopsis*

The identification of the 77 bp sequence as the MGSE harboring region (refer to 5.2.5) has led to the conclusion that the MGSE is present in commonly used *Arabidopsis* ecotypes such as Landsberg erecta, Columbia and C24 after comparing their corresponding sequences. However, whether this can be generalized to all ecotypes of *Arabidopsis* is still a question. The answer to this question may be important for predicting the possible function of the MGSE in the genome.

Other ecotypes (some are rare) were then chosen to do a further survey for the distribution of the MGSE among different ecotypes. Seeds from ecotypes Aua/Rhon, Benshein, Cape Verde Islands, Dijon-G, Estland, Greenville, Muhlem, Nossen, S96 and Wassilewskija (WS) (purchased from LEHLE SEEDS, USA) were sown on soil for germination and the seedlings were grown to two-leaf stage for DNA preparation. Genomic DNA was prepared from the leaf tissue of each ecotype by the mini-preparation method described in 2.1.2. These DNA samples were then used as templates for PCR amplification of the approximately 490 bp fragments, which correspond to the region covering the 318 bp MGSE containing sequence and the 165

bp MITE-like sequence in ET253, using primer pairs ET-TL13-1 (refer to 5.2.5) and ET-BP50 (5'- GCCCAAGCTTCCATGGCTGTGAATGCTAACCA-3').

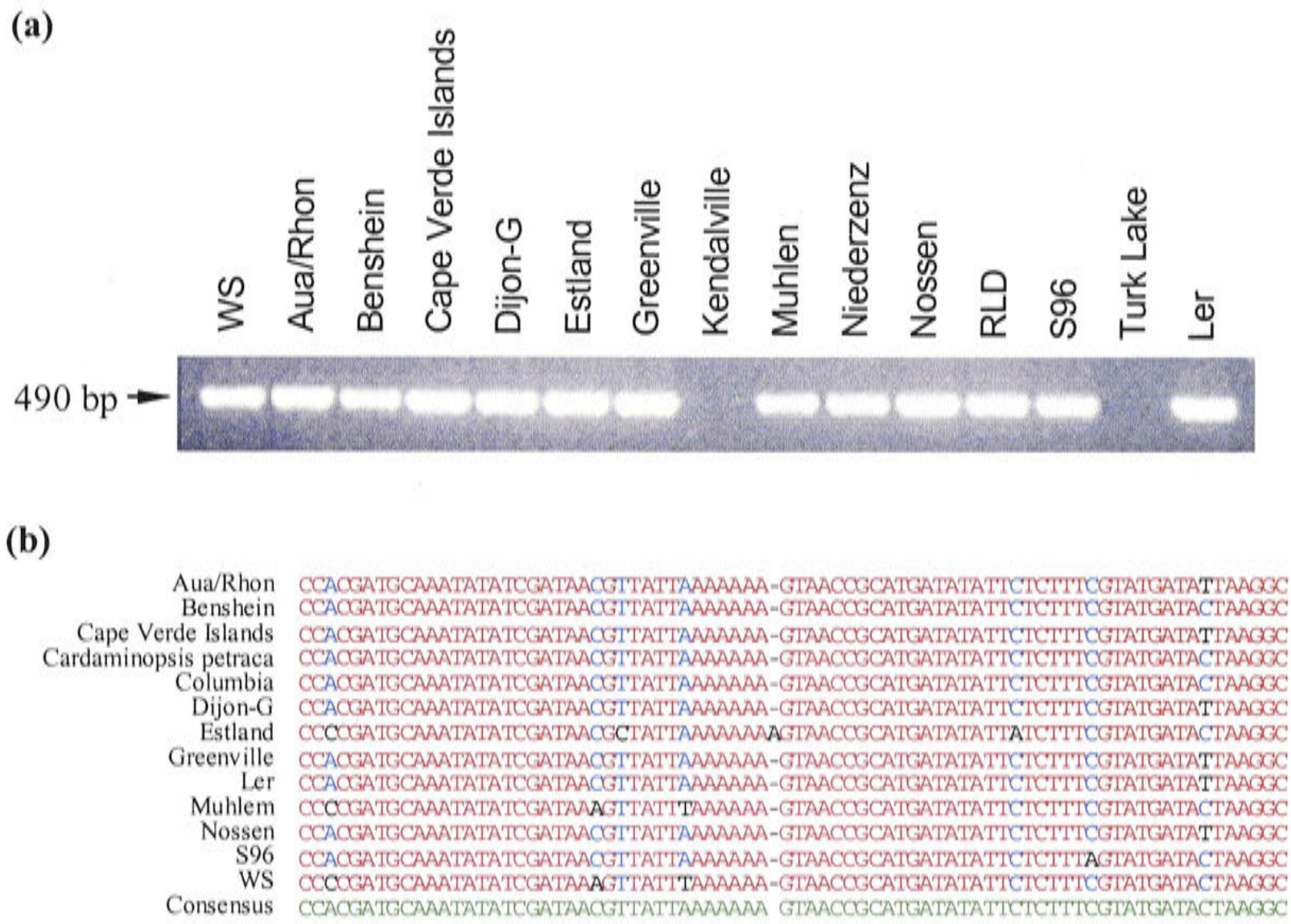


Figure 5.14 Distribution of the MGSE among different ecotypes of *Arabidopsis*.

(a) PCR amplification of the region covering the MGSE and the MITE-like transposable element in different ecotypes.

(b) Alignment of the 77bp sequence with MGSE activity from different ecotypes. Conserved nucleotides are in red color, and identical nucleotides in half or more are in blue color.

As shown in Figure 5.14a, the 490 bp fragment was amplified from ecotypes Aua/Rhon, Benshein, Cape Verde Islands, Dijon-G, Estland, Greenville, Muhlem, Nossen, S96 and Wassilewskija (WS) but not from ecotypes Kendalville and Turk Lake. The 490 bp PCR product from different ecotypes were then sequenced and the 77 bp MGSE-bearing sequence was compared. Alignment of the 77bp sequence from

different ecotypes shows that the MGSE-bearing sequences are very conserved (Figure 5.14b). In fact, the whole 490 bp fragment, including the MITE-like sequence (refer to 4.2.3), from all these ecotypes are also conserved except for some single nucleotide changes (data not shown). This eliminated the possibility that the independent MGSE is the result of an ecotype specific genomic insertion, deletion or rearrangement.

5.3 Discussion

5.3.1 The MGSE is a cis-acting element

As described in Chapter 1, the main difference between gene trap and enhancer trap is that the gene trap detects the defined transcribed DNA region whereas enhancer trap detects the cis-acting DNA element in a broad range. As a result, the frequency of reporter gene expression detected in gene trap lines is much lower than that detected in enhancer trap lines. However, once a gene trap line with particular reporter gene expression pattern is obtained, the procedure for the identification of the responsible gene is simple. In contrast, although many reporter gene expression patterns can be easily obtained from enhancer trap lines, the localization of the corresponding enhancer is not a trivial task due to the several possibilities related to the enhancer action as discussed in Chapter 4. A recent discovery showing that enhancer can be functional *in trans* when mediated by certain transcription factors (Mahmoudi *et al.*, 2002) makes it even more complicated to deal with an enhancer. In the case of *in trans* activation in certain enhancer trap lines, the responsible enhancers will not be possible to locate because they are not tagged.

In *Arabidopsis*, the situation involving enhancers acting through *in trans* activation in a natural environment has not been reported. Therefore, we assumed that the GUS expression pattern in ET253 should still represent cis-activation. Based on

this assumption, we tried to analyze the DsE flanking sequence by testing different DsE flanking regions from ET253 in wild type *Arabidopsis* and successfully located the region harboring the MGSE.

5.3.2 The interaction between the 35S minimal promoter and the 35S full promoter and dealing with the 35S effects

To locate the MGSE, we used the strategy to take different segments of the upstream and downstream regions of the DsE insertion site in ET253 by fusing them to the *gusA* gene and test them in wild type *Arabidopsis*. In this way, we were able to find out which segment contains the regulatory element that can reproduce the megagametophyt-specific GUS expression pattern as seen in the original enhancer trap line ET253.

pWY-F68 is the construct that can lead to the GUS expression in the megagametophyte of its transformed plants as presented in 5.2.1. However, when the 35S minimal promoter is introduced into the pCAMBIA1201 background to control the *gusA* gene, strong GUS expression was unexpectedly detected in different tissues of the transgenic plants except for the ovules, as demonstrated with the two constructs pWY-I13 and pWY-K35.1 in 5.2.3. Although some of the expression patterns could in theory be controlled by the genomic enhancers around the T-DNA insertion site(s), the majority may represent the internal effect from the full 35S promoter within the T-DNA. This is consistent with a recent report showing that the 35S promoter, which is a polar (unidirectional) promoter itself, can be bidirectionalized by adding a 35S minimal promoter at its 5' end in opposite orientation (Xie *et al.*, 2001). The bidirectionalization can also be extended to other polar promoters like the peanut chlorotic streak caulimovirus (PCISV) full-length transcript promoter, the senescence-specific SAG12 promoter and the OPR1 inducible promoter using either a homogeneous or a heterogeneous minimal promoter (Xie *et al.*, 2001). Now, it is known that the bidirectionalization is due to the cis-activation of the minimal promoter by the UPEs in

the polar promoter as described in 5.2.3. Furthermore, the results from the analysis of the transgenic plants from pWY-F68, pWY-J26.2 and pWY-J47.3 also showed that this kind of cis-activation can not be blocked by the spacing between the polar promoter and the compatible minimal promoter, at least not within the distance of 2 kb as in the case of pWY-F68. By combining all the above information, we realized that extreme care needs to be taken to avoid the interference of the expression pattern under the control of an enhancer to be tested by some internal cis-activation of the reporter gene when an enhancer-testing vector is designed.

5.3.3 The advantages of pWY-K105.1 and pWY-O93.4

In this study, two new vectors pWY-K105.1 and pWY-O93.4 for testing enhancer activity of certain DNA fragment have been designed and constructed. When such vectors were first designed, focus was mainly on the following three points: 1. No intrinsic enhancer activity carried by the vector itself that may interfere with the enhancer activity to be tested; 2. Easy for the cloning of candidate DNA fragments; 3. High efficiency in obtaining transgenic plants. pCAMBIA vectors such as pCAMBIA1201 and pCAMBIA1281Z (www.cambia.org) have been shown to be good binary vectors for *Agrobacterium*-mediated plant transformation including *Arabidopsis*. Modification of pCAMBIA1201 was therefore chosen so that the derived vectors can meet the criteria for the above three points.

For the first point, it is necessary to remove the 35S promoter driving the selectable marker gene for plant selection and replace it with a promoter that will not interact with the 35S minimal promoter. pWY-K35.1 (Figure 5.5a) is a derivative of pCAMBIA1201 as described in 5.2.3 . It contains 35S minimal promoter fused to the *gusA* gene with the catalase intron and the multiple cloning sites (MCS) from pUC18 that can meet well the requirement for the second point. It still keeps the wide-host-range origin of replication from plasmid pVS1, which is extremely stable in the absence

of selection for replication in *Agrobacterium*, and the pBR322 ORIGIN (pMB9-type) to allow high-yielding DNA preparations in E. coli (Hajdukiewicz *et al.*, 1994). This intermediate plasmid was therefore used as a backbone for the construction of the new vector pWY-K105.1 by simply replacing the 35S promoter-*hgy* cassette with the 1'-*bar* cassette in opposite orientation (as shown in Figure 5.5). The reason for placing the 1'-*bar* cassette in this orientation is that the "so called" 1' promoter is actually still the 1'-2' dual promoter (Velten *et al.*, 1984). This arrangement can prevent the effect from the 2' promoter activity, because the 2' effect is evident when the 1'-*bar* cassette is placed in the same orientation as the 35S promoter-*hgy* cassette (data not shown). As described in 5.2.4, the pWY-K105.1 transformed *Arabidopsis* plants did not display any visible GUS expression. The transformation efficiency of pWY-K105.1 is also high, partly due to the *bar*-gene-based effective selection system, making it easy to generate enough transgenic plants for a broad range of analysis. All these features suggest that pWY-K105.1 can serve as a good vector for testing enhancer function.

pWY-O93.4 is an alternative to pWY-K105.1 by replacing the *gusA* gene with the EGFP gene, giving more flexibility to this enhancer-testing vector family, especially when localization of certain enhancer activity is preferred in live tissues. It is believed that these two enhancer-testing vectors could be widely used in the scientific community for examining enhancers with different tissue specificity in the future, due to the increasing use of enhancer trapping systems in plants in searching for new genes and regulatory elements.

5.3.4 Defining the core sequence for the MGSE

Since the GUS gene expression under the control of the MGSE is restricted to a few cells in the embryo sac, which are very difficult to access, the determination of this type of GUS expression from transgenic lines of different constructs (such as those used in deletion tests) is not an easy task. At present, the expression can only be monitored

qualitatively by histochemical GUS staining, rather than quantitatively as the GUS assay suitable for larger quantity of tissues such as vegetative tissues (Jefferson, 1987). Moreover, even the qualitative determination of the GUS expression in the embryo sac is often affected by the following two factors: ovule developmental stage and the level of GUS expression.

In Chapter 3, it has already been shown that the intensity of GUS expression in the ovules of ET253 varies among different floral stages, with the peak at around stage 14. In addition, the intensity of GUS expression may also vary among ovules within one gynoecium due to the variation in ovule developmental stages although one developmental stage normally predominates within most gynoecia (Christensen *et al.*, 1997). On the other hand, the level of GUS expression is also associated with the copy numbers of T-DNA insertions carrying the MGSE-mp-*gusA* expression unit, as shown in 5.2.8. Some lines showed no observable GUS expression although the T-DNA is integrated in the genome. Therefore, it is very important to examine enough numbers of independent transgenic lines from one construct to determine the overall performance of the construct. Furthermore, it is also important to examine all the flowers (buds) on more than one inflorescence from a plant so that all developmental stages can be covered. Based on these considerations, the following steps were set up for determining whether a sequence possesses the MGSE activity when deletion test was conducted: 1. Generating a reasonable number of transgenic lines from each deletion construct (normally ten lines should be enough); 2. Staining at least two inflorescences from each individual line with GUS staining solution under standard condition as described in 2.8; 3. Observing all flowers (buds) covering different developmental stages. The flowers at around stage 14 were mainly focused on due to better GUS staining visually at that stage in general. The result of the observation is a simple yes or no answer. If all transgenic lines from one construct show no observable GUS expression, the sequence being tested is counted as carrying no MGSE activity. However, even if only one line shows GUS activity, the sequence being tested is count

as having MGSE activity. In the real deletion experiments, the majority of transgenic lines carrying deletion construct containing different sequences with MGSE activity displayed GUS expression (Table 5.3).

By using the above method, the sequence with the MGSE activity was narrowed down to 77 bp by 5' and 3' deletion tests as presented in 5.2.5. However, in an attempt to further define the MGSE core sequence within the 77 bp fragment by scanning for MGSE activity with five 30 bp units and one 27 bp unit with 20 bp overlapping sequence between the adjacent units (Figure 5.8), it was failed to obtain the finer sequence of the MGSE. This at least suggested that the MGSE core sequence is either bigger than 30bp or arranged in separate domains. A fine mutation test (Ito *et al.*, 1998) of the 77 bp sequence may be needed to determine the core sequence of the MGSE. However, the result showed that the 77 bp sequence is efficient in controlling specific gene expression in the embryo sac of *Arabidopsis*.

5.3.5 Basic features of the MGSE

By studying the 77 bp MGSE sequence, it is demostrated that the MGSE possesses several clear features. Some of the features could be very useful when it comes to the stage to practically use the MGSE in controlling certain gene expression in the process of seed formation.

Firstly, the MGSE has a single copy in *Arabidopsis* genome. Based on the BLAST search and Southern analysis, we know that the 77 bp sequence covering the MGSE is unique in the *Arabidopsis* genome. Considering that the MITE-like repetitive sequence directly downstream of the 77 bp sequence (Figure 5.7) has over 30 copies in the *Arabidopsis* genome as described in 4.2.3, one will naturally think that the MGSE is not part of this MITE-like repetitive element. Sequence alignment of the MITE-like repetitive element from different chromosomes also confirmed that there is a clear

boundary between the 77 bp MGSE-bearing sequence its directly downstream repetitive element.

Secondly, the MGSE sequence is very conserved among different ecotypes of *Arabidopsis thaliana*. As described In 4.2.10, the region covering the 318 bp MGSE-containing sequence and the MITE-like repetitive element is conserved among the most commonly used ecotypes such as Landsberg erecta, Columbia, C24 and Wassilewskija, and other ecotypes like Aua/Rhon, Benshein, Cape Verde Islands, Dijon-G, Estland, Greenville, Muhlem, Nossen and S96. However, Kendalville and Turk Lake may be different because it was failed to amplify the same region by PCR from the two ecotypes. This could be due to certain genomic rearrangements and a further study on these two ecotypes may lead to interesting results related to the evolution of ecotypes of *Arabidopsis thaliana*. In addition, some close relatives of *Arabidopsis thaliana*, such as *Arabidopsis petraea* (formerly in the genus *Cardaminopsis*) (Koch *et al.*, 1999; Koch *et al.*, 2000), *Arabidopsis griffithiana* (Price and Palmer, 1994; Koch *et al.*, 1999) and *Arabidopsis pumila* (Price and Palmer, 1994), were also included in the survey. We were able to PCR amplify the same region covering the 318 bp MGSE-bearing sequence and the MITE-like repetitive element from *Arabidopsis petraea* with the sequence same as Columbia (its 77 bp MGSE bearing sequence is shown in Figure 5.14b), but failed to obtain any PCR amplification from *Arabidopsis griffithiana* and *Arabidopsis pumila* (data not shown). This at least confirmed that the MGSE is present in some of the close relatives of *Arabidopsis thaliana* but further study is needed to elucidate the underlying relationships.

Thirdly, the MGSE is position independent. In the original enhancer trap line ET253, the 77 bp MGSE-containing sequence is separated from the mp-*gusA* unit by 221 bp Ds sequence. However, when the 77bp MGSE-containing sequence was placed directly upstream of the 35S minimal promoter fused to the *gusA* gene in all the 5' deletion constructs as presented in 5.2.5, it still showed good MGSE activity. This

means that the MGSE is not sensitive to its position change within certain range in the upstream region of the gene it interacts with. However, we have not examined whether the MGSE can work efficiently over a long range, e.g. several kb away, or work from downstream of the stimulated promoter. If it can function in this manner, then the use of the MGSE in controlling specific gene expression would be more flexible when constructs are designed.

Fourthly, the MGSE is orientation independent. This orientation dependent feature was determined based on the 77 bp sequence. It is however contradicted by the result from the 318 bp sequence that showed no MGSE activity when placed in reverse orientation (Figure 5.10) as noted in 5.2.7. According to the above discussion, it is unlikely to be the positional effect that stopped the MGSE activity. It is therefore speculated that there may be an insulator upstream of the 77 bp sequence within the 318 bp fragment. An insulator, or boundary element, is a class of regulatory sequences that play a role in limiting enhancer activity and thereby establishing independent functional domains within the chromosome (Geyer, 1997; Zhan *et al.*, 2001). Enhancers normally stimulate transcription in an orientation-independent manner when assayed on circular plasmids. However, enhancers may have an orientation preference in a chromosomal context due to the presence of certain boundary elements (Wei and Brennan, 2000). So far, most experiments designed to test the orientation preferences of enhancers have used relatively small or core enhancer elements, which may not be representative of native chromosomal enhancers. Considering this, enhancers might be expected to display an orientation preference when assayed using a region of DNA sufficiently large to mimic the arrangement of the natural chromosome. The MGSE may also be arranged in the same manner by the presence of certain insulator sequence on one side and thus make the MGSE activity with polarity. If this is true, the MGSE would then only control a downstream gene, such as a gene beyond the *ASL41* gene, in the real genomic context. This speculation needs to be further examined by testing whether the upstream region of the 77 bp sequence has the insulator activity. A polar

MGSE could be more useful than its orientation independent core sequence when a precise control of gene expression in the megagametophyte is desired.

Finally, the MGSE may display higher activity with its tandem repeats. Although this assertion has not been backed by quantitative measurement, the results from 5.2.8 still suggested that the level of the MGSE activity could possibly be adjusted by using different numbers of the element in a tandem array. This feature could be used for overexpression of certain gene in embryo sac for a loss of function test.

5.3.6 The endogenous role of the MGSE in the *Arabidopsis* genome remains a question

We have cloned and characterized the MGSE from *Arabidopsis* using enhancer trap. The results show that the sequence containing the MGSE is located at quite a distance from the nearby genes (Figure5.7). Without using enhancer trap, it may never have been uncovered. Where does this enhancer originate from and what could its functions be in the genome? There are only three genes in the 11.5 kb MGSE flanking region: two ATP9a-like peroxidase genes *atpH21-1* and *atpH21-2* as tandem arrays and the *ASL41* gene (referred to as genes F19F18.10, F19F18.20 and F19F18.30, respectively, in BAC clone F19F18). The MGSE is in the middle of the region between the *atpH21-1* gene and the *ASL41* gene. According to the orientation of these two genes, the MGSE is in the downstream region of both genes. This fact eliminated the possibility that the MGSE is part of a promoter of a gene. The region between the *atpH21-1* gene and the *ASL41* gene is 3.4 kb but it has no unambiguous open reading frame (ORF), also suggesting that the MGSE be in an intergenic region. In addition, The MGSE does not belong to the direct downstream MITE-like repetitive element (Figure 5.7).

The single copy nature and conservation of the MGSE among different ecotypes indicate that the location of this element is not the result of a sudden genomic

change (such as transposon insertion) in a particular ecotype. It therefore suggests that the MGSE may have a defined function in *Arabidopsis* genome rather than just be a cryptic enhancer. One possibility is that it can activate a gene beyond the three nearby genes described here in a remote fashion, for example, activation *in trans* (Chen *et al.*, 2002), so that the targeted gene expression can be restricted to the megagametophyte and be included in the pathways involved in the embryo sac development and early embryo formation. In this case, there is a need to extend the search range for potential candidate genes.

CHAPTER 6

EXPLORING THE USE OF THE MGSE ELEMENT

6.1 Introduction

It has been shown in Chapter 5 that the MGSE element is able to activate CaMV 35S minimal promoter in driving reporter gene expression (both GUS and GFP genes) specifically in the megagametophyte of *Arabidopsis*. The next question is: can it be used as a tool to control functional gene expression in female gametophyte? To answer this question, an appropriate gene has to be selected. Currently, very few genes expressed in the female gametophyte have been isolated (Drews *et al.*, 1998) which gives almost no choice to test genes with known function in the female gametophyte. However, to address the above question, the gene coding for diphtheria toxin chain A (DTA) from *Corynebacterium diphtheriae* (Greenfield *et al.*, 1983) could be very useful in testing the tissue specificity of the MGSE. This gene has been successfully used for genetic ablation in mammalian (Breitman *et al.*, 1987; Palmiter *et al.*, 1987) and *Drosophila* (Bellen *et al.*, 1992), and has also been shown to be active in plants (Thorsness *et al.*, 1993; van der Geest *et al.*, 1995; Day *et al.*, 1995; Nilsson *et al.*, 1998).

The DT toxin contains two functional domains: chain A, which carries the active site for ADP-ribosylation of the enlongation factor 2 (EF2) that inhibits all protein synthesis, and chain B, which is required for the binding of the toxin to cells and the entry of chain A into the cytosolic compartment (Greenfield *et al.*, 1983). Without chain B and other extracellular targeting sequences, DTA can not be transported across the plasma membrane and remains only in the cells where it is expressed. Cells expressing DTA are not able to synthesize proteins and divide, and will therefore soon die. This property of DTA has been used for targeted ablation of spatially differentiated

cells (Thorsness *et al.*, 1993; Day *et al.*, 1995; Tsugeki and Fedoroff, 1999) and cells in particular developmental stages (van der Geest *et al.*, 1995) in plant when an appropriate (spatial or temporal) promoter was used. If the DTA gene is placed under the control of the MGSE, the specificity of the MGSE should determine the location of the DTA expression. The use of tissue-specific promoters allows the transcription to be restricted to a defined subset of cells. However, if the gene product to be expressed is toxic to the organism, any leakage of the toxin will make it difficult to establish stable transgenic lines carrying the chimerical gene.

To carry out such cell ablation test, the GAL4-UAS transactivation system (Ma *et al.*, 1988) was chosen. This system is based on the yeast Gal4 transcriptional activator, which consists of a DNA-binding domain that binds to the upstream activation sequence (UAS) in a promoter and an activation domain that interacts with the components of the transcriptional machinery assembling on the promoter. A hybrid transcriptional activator Gal4-VP16 with the Gal4 DNA-binding domain fused to the acidic activation domain of the herpes simplex virus VP16 protein (Triezenberg *et al.*, 1988) has shown to be even more potent transcriptional activator in mammalian cells (Sadowski *et al.*, 1988). This Gal4-VP16 hybrid transcriptional activator also works well in maize, tobacco and *Arabidopsis* (Goff *et al.*, 1991; Aoyama and Chua, 1997). For targeted gene expression, a system based on GAL4-UAS transactivation has been developed which separates the targeted gene under the control of the Gal4 binding UAS from its transcriptional activator in two distinct transgenic lines. In essence, in one line (target line) the target gene remains silent in the absence of its activator and in the second line (activator line) the activator protein is present but has no target gene to activate. Expression of the target gene is achieved by crossing the target line with the activator line. In this way the target gene will be expressed only in those cells of the progeny in which the transcription activator is expressed. This system has been successfully applied on insect (Brand and Perrimon, 1993), animal (Scheer and Camnos-Ortega, 1999) and plant (Moore *et al.*, 1998) for regulated gene expression.

This chapter presents an attempt to use the DTA toxin in combination with the Gal4-VP16 transactivation system as a way to achieve specific cell ablation while testing the specificity of the MGSE. *Arabidopsis* lines expressing the hybrid Gal4-VP16 transcription activator under the control of the MGSE plus a minimal CaMV 35S promoter were generated. By crossing these lines with an *Arabidopsis* line harbouring the *DTA* gene under the control of Gal4 upstream activator sequences (a gift from Dr Jim Haseloff, University of Cambridge), very specific ablation of embryos in the hybrid seeds was achieved. More importantly, these specific activator lines could be used to control the expression of other functional genes by crossing with other target lines containing the gene of interest.

6.2 Results

6.2.1 The activator line

6.2.1.1 Design and construction of activator construct

The establishment of good activator lines in *Arabidopsis* depends largely on the specificity of the promoter controlling the expression of the transcription activator. To achieve tissue specific cell ablation using a transactivation system, it is necessary to assemble a specific promoter to restrict the expression of the transactivator in certain cells. In addition, the expression pattern of the transactivator needs to be examined before a cross with target line is conducted. For this purpose, a construct consisting of a Gal4/VP16 expression unit under the control of the MGSE and an UAS-GUS reporting unit under the control of Gal4 specific UAS in one T-DNA was designed. This design makes it convenient to check the expression pattern of Gal4/VP16 on reporter gene in T1 transgenic activator lines without crossing.

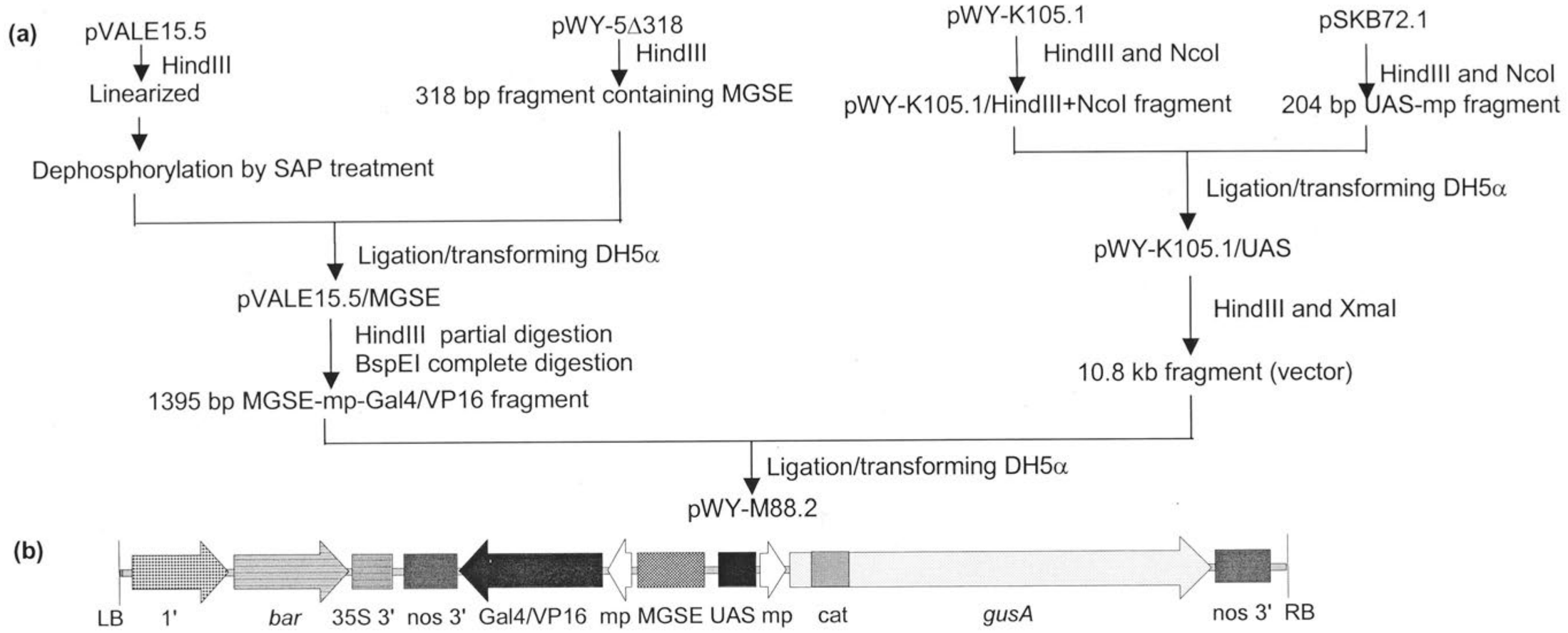


Figure 6.1 Construction of pWY-M88.2

(a) Procedure for the construction of pWY-M88.2.

(b) Schmetic map of the T-DNA region of pWY-M88.2. 1', 1' promoter; bar, Basta resistance gene; 35S 3', CaMV 35S gene terminator; nos 3', NOS gene terminator; Gal4/VP16, coding sequence for Gal4 DNA binding domain and VP16 activation dmain; mp, CaMV 35S minimal promoter; cat, catalase intron; *gusA*, β-glucuronidase gene; LB, T-DNA left border; RB, T-DNA right border.

Figure 6.1a is a brief flow chart of the procedure in constructing such a construct. For assembling the Gal4/VP16 expression unit, a Gal4/VP16 coding sequence driven by a 35S minimal promoter (+8 to -46) from plasmid pVALE15.5 (from Dr Valasubramanian Ramaiah, CAMBIA, unpublished data) was placed under the direct control of the 318 bp fragment with MGSE activity from pWY-5Δ318 to generate an intermediate plasmid pVALE15.5/MGSE. Since the 318 bp fragment in reverse orientation does not show megagametophyte-specific enhancer activity (refer to Chapter 5), the direction of the 318 bp fragment was placed in the original orientation, i.e. the same orientation as that in line ET253 for controlling the GUS expression. For assembling the UAS-GUS reporting unit, a fragment containing 6×UAS placed upstream of a 35S minimal promoter (UAS-mp) from plasmid pSKB72.1 (from Sri Koerniati, CAMBIA, unpublished data) was inserted upstream of the *gusA* gene in pWY-K105.1 to form another intermediate plasmid pWY-K105.1/UAS. To assemble the two units into one T-DNA, the MGSE-mp-Gal4/VP16 fragment from pVALE15.5/MGSE was cloned into pWY-K105.1/UAS to generate the final plasmid pWY-M88.2.

Detailed arrangement of the T-DNA region in pWY-M88.2 is shown in Figure 6.1b. In this T-DNA, the Gal4/VP16 expression unit and the UAS-GUS reporting unit are in opposite directions such that the GUS gene expression can only occur after being transactivated by the Gal4/VP16 protein expressed under the control of the MGSE. The GUS gene expression should therefore indicate where the Gal4/VP16 is expressed.

6.2.1.2 Generation of activator lines

Activator lines were generated by transforming the activator construct pWY-M88.2 into wild type *Arabidopsis*. Transgenic plants were selected against Basta and from which

candidate activator lines for subsequent crossing with the UAS-DTA line were selected by their GUS expression pattern. In this experiment, *Arabidopsis* ecotype C24 was used as recipient and the transformation through floral dip and the Basta selection of transgenic plants were the same as described in 2.7.2 and 5.2.4. The reasons for choosing ecotype C24, instead of ecotype Landsberg erecta as previously did for the identification of the MGSE, to generate activator lines were mainly subject to the following two aspects: 1. the UAS-DTA line (refer to 6.2.4.2), which will be used to cross with the activator lines for cell ablation test, is in C24 background. Thus, making the activator lines in C24 background will facilitate the follow up crosses when compatibility is concerned; 2. Landsberg erecta has been shown to be a poor recipient for transformation according to Clough and Bent (1998) and our previous deletion experiments (Chapter 5), whereas C24 is very close to Columbia, which consistently showed good transformation frequencies (Mengiste *et al.*, 1997; Clough and Bent, 1998b; Ye *et al.*, 1999). It was hoped that C24 could make it easier to generate more transgenic lines and thus would help for the analysis of the behavior of the Gal4/VP16 expression when large sample number was needed.

The transformation result showed that C24 is much better than Landsberg erecta in producing transgenic plants. From ten EHA105/pWY-M88.2 dipped C24 plants, approximately 2 ml of seeds were obtained. Thirty-eight Basta-resistant plants were obtained from half of the seeds selected by Basta. These plants were candidates to be selected and used as activator lines

6.2.1.3 GUS expression patterns of the pWY-M88.2 transformed lines (activator lines)

When reporter genes such as *gusA* or *GFP* were placed under the direct control of the MGSE, their expression was restricted to the embryo sac (refer to Chapter 5). To examine the patterns of GUS expression in the activator lines harbouring *gusA* gene under indirect control of the MGSE through transactivation, GUS staining of different

tissues from the Basta-resistant T1 plants was carried out. Twenty-five plants were randomly selected as representatives from the thirty-eight Basta-resistant plants obtained for GUS pattern determination. Initially, it was expected that the pWY-M88.2 transformed lines should still have GUS expression in the embryo sac if there is no cis-activation from the genomic region nearby the T-DNA insertion(s) in each individual transgenic plant. Surprisingly, however, diverse GUS expression patterns were observed from different transgenic plants.

Table 6.1 GUS expression patterns of the pWY-M88.2 transformed T1 lines

T1 line	GUS expressed tissues	T1 line	GUS expressed tissues
1	M, Ca, Sty, Se	14*	M
2	M, N, Sty, L	15	-----
3	-----	16	A
4	A, N, Sty, R	17	A, R, N, Sty, L
5	-----	18	A, L
6	A, N, L, Sty, Ca	19	A, Fila, Se, L
7	A, L, R, Sty	20*	M
8	M, A, R, Sty, L, Ca	21*	M
9	M, Se, Sty, Ca, L	22	A, L, Se, Sty, Sti, Stem
10	-----	23	M, R, Se
11	A	24	M, R, Sty, Se
12	M, Sty, Se	25	A, R
13	R, Se		

Abbreviations: M, megagametophyte; A, anther; L, leaf; Sty, style; Sti, stigma; Se, Sepal; R, receptacle; N, nod; Ca, carpel; Fila, filament.

* Megagametophyte-specific expression

The GUS staining results from the twenty-five T1 lines are summarized in Table 6.1. As can be seen in this table, twenty-one lines (84%) display GUS expression in at least one tissue type and four lines (16%) show no detectable GUS expression at all. Among the lines with GUS expression, three types of GUS expression are revealed if GUS expression in the megagametophyte is concerned. Three lines (lines 14, 20 and 21) show GUS expression only in the embryo sac as expected. However, seven lines (lines 1, 2, 8, 9, 12, 23 and 24) show GUS expression not only in the embryo sac, but also in other tissues such as leaf and other floral tissues, including sepal, style, receptacle and anther. In contrast, the other 11 lines (lines 4, 6, 7, 11, 13, 16, 17, 18, 19, 22 and 25) do not have detectable GUS expression in the embryo sac but show GUS expression in other tissues as described above.

Despite the diversity of the GUS expression in the pWY-M88.2, nearly half of the lines showing GUS expression in the embryo sac clearly demonstrated that transcriptional activation has occurred at least in these lines. This makes the follow up transactivation of the DTA gene by crossing feasible.

6.2.1.4 Selection of the activator line

The GUS expression results obtained from the pWY-M88.2 transgenic lines are very different from that obtained from the plants receiving plasmid containing *gusA* gene under the direct control of the MGSE, which showed only megagametophyte-specific GUS expression. Therefore, it is necessary to choose an appropriate line from these transgenic lines for being used as activator line.

Although the GUS expression patterns of different pWY-M88.2 transgenic lines are quite diversified, a basic pattern is apparent. That is, the GUS expression is mainly restricted to style (44%), leaf (36%), receptacle (32%) and sepal (32%) (Table 6.1), where normally show intensive staining. Although many lines also show blue staining in

anthers, the staining in most of those lines is weak that may still be due to artifact and therefore it is not considered as part of the basic pattern. This basic pattern may reflect the background expression pattern from the Gal4 system because the Gal4 DNA-binding domain itself has been shown to stimulate report gene expression to some extend in maize and tobacco (Goff *et al.*, 1991; Schlappi *et al.*, 1996). It is therefore reasonable to speculate that the Gal4 DNA-binding domain can activate reporter gene expression in certain cell types in *Arabidopsis*. However, the GUS expression in other tissue types such as stigma, carpel, nod and filament is possibly due to the influence of the nearby regulatory elements when a T-DNA was inserted in certain genomic regions.

To use a pWY-M88.2 transformed line as an activator to cross with any target lines, the lines with megagametophyte-specific GUS expression are certainly the first choice. Among three such lines (see 6.2.1.3), line 14 showed the strongest GUS expression in the embryo sac (data not shown). It was therefore chosen as the activator line for the subsequent cross with the target line.

6.2.2 The UAS-DTA line as target line

The UAS-DTA line is an *Arabidopsis* line generated in Dr Jim Haseloff's lab (Bougourd *et al.*, 2000). This line is one of the very few lines survived after the transformation of an UAS-DTA construct into *Arabidopsis*. The UAS-DTA construct contains a *DTA* gene controlled by an artificial promoter consisting of multiple copies of the UAS sequence recognized by Gal4 and a minimal CaMV 35S promoter (Figure 6.2). In the absence of the Gal4/VP16 protein, the *DTA* gene should keep silent theoretically. However, the reality is that it is difficult to obtain transgenic plants from this construct. The main reason for the very low transformation efficiency is probably that any leakage of the DTA expression, activated by some native regulatory elements in vicinity for instance, is lethal to its transformants (Jim Haseloff, personal communication).

This UAS-DTA line is completely normal and fertile. It can be propagated easily by selfing. When it was crossed with wild type C24, fertile F1 seeds were obtained without any problem (Table 6.2 and Figure 6.3). This means that the UAS-DTA line has no leaking DTA expression. On the other hand, It has been shown that the DTA gene in this line can be activated by genetic crossing to Gal4 lines (Bougourd *etal.*, 2000).

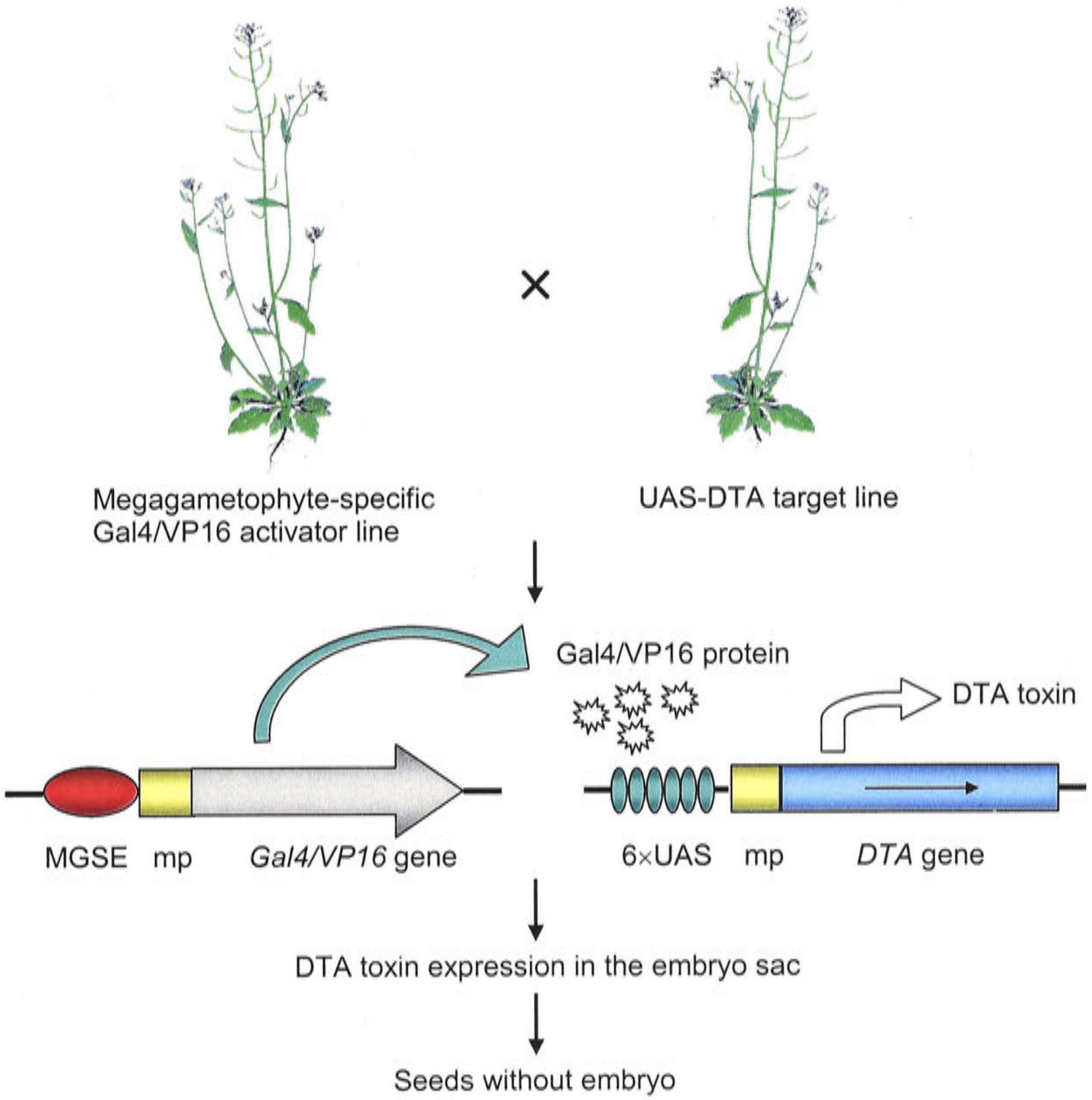


Figure 6.2 Scheme for the embryo sac specific activation of the *DTA* gene in *Arabidopsis*.

In a two-part system based on Gal4/VP16 transactivation, the Gal4/VP16 gene and the *DTA* gene are placed in two separate lines. The MGSE element is used to control the specific expression of Gal4/VP16 in the embryo sac in the

Gal4/VP16 activator line. In the UAS-DTA target line, the DTA gene is under the control of the Gal4 binding sites (UAS) and it is silent in the absence of Gal4/VP16 protein. Embryo sac specific DTA expression is achieved by crossing the two lines, which brings the Gal4/VP16 and the Gal4 UAS in close proximity. As a result, the embryos of the hybrid seeds are specifically ablated.

6.2.3 Generation of embryo-defective F1 seeds through specific transactivation of the DTA toxin in the embryo sac

To activate the *DTA* gene, the Gal4/VP16 transcription factor and the *DTA* gene under the control of the UAS sequence need to be brought together by fertilization. Figure 6.2 is a brief illustration of such an approach. Reciprocal crosses between the activator line (the pWY-M88.2 transformed line 14) and the target line (the UAS-DTA line) were carried out for testing the transactivation of the *DTA* gene when either the activator line or the target line was used as female parent.

As described in 6.2.2, the UAS-DTA line has normal seed setting if pollinated with pollen from C24 and *vice versa*. However, when the UAS-DTA line was pollinated with pollen from the activator line, abortive seeds were produced (Figure 6.3). The abortive seeds are shriveled but have seed coats in normal size, indicating that the integument development is normal. This type of abortive seeds shares some common features with the abortive seeds obtained from those embryo-defective mutants with embryo arrested at early stages of seed development (Meinke, 1994). After crossing, the UAS-DTA line has normal silique development despite the defect in the seed attached to it (Figure 6.4). These results therefore strongly suggest that embryo-specific ablation occurred through the transactivation of the *DTA* gene after cross. In comparison, when the activator line was pollinated with pollen from the UAS-DTA line, same result was obtained. In most crosses, 100% abortive seeds were generated (Table 6.2). The small amount of normal seeds from some crosses, such as UAS-

DTA×14-5 and 14-3-2×UAS-DTA, is probably due to contamination in the emasculation and pollination processes.

Figure 6.3 Different F1 seeds from the crosses between the UAS-DTA line and the plants with or without embryo sac specifically expressed Gal4/VP16.

(**A**) Opened siliques produced by an UAS-DTA plant pollinated with eiher pollens from a C24 wild type plant showing normal seeds (top), or pollens from an MGSE-Gal4/VP16 plant (the pWY-M88.2 transformed line 14 in C24 background) showing abortive seeds (bottom).

(**B**) and (**C**) The close-up view of the mature normal seeds and the abortive shrivel seeds produced by the above two crosses. Note that this type of abortive seed has normal seed coat development but the embryo is ablated.

Table 6.2 Seed setting after reciprocal cross between the pWY-M88.2 transformed line 14 and the UAS-DTA line

Crosses (E×Γ)	No. of siliques[a]	No. of good seeds[b]	No. of abortive seeds[c]
Control			
1. UAS-DTA×C24	5	122	4
2. C24× UAS-DTA	14	289	3
Experiment			
3. UAS-DTA×14-3-4	2	0	33
4. UAS-DTA×14-3-9	9	0	187
5. UAS-DTA×14-5	13	7	305
6. 14-3-1× UAS-DTA	7	0	99
7. 14-3-2× UAS-DTA	6	4	138
8. 14-3-4× UAS-DTA	7	0	203
9. 14-5× UAS-DTA	7	0	167

[a] Number of well developed siliques after crossing. Siliques failed to develop after crossing are not included.

[b] Number of seeds collected from the well developed siliques.

[c] Number of aborted seeds from the well developed siliques.

Note: 14-3-1, 14-3-2, 14-3-4, 14-3-9 and 14-5 represent different plants of the pWY-M88.2 transformed line 14. The number after the first '-' stands for the T_2 plant number and the number after the second '-' stands for the T_3 plants number.

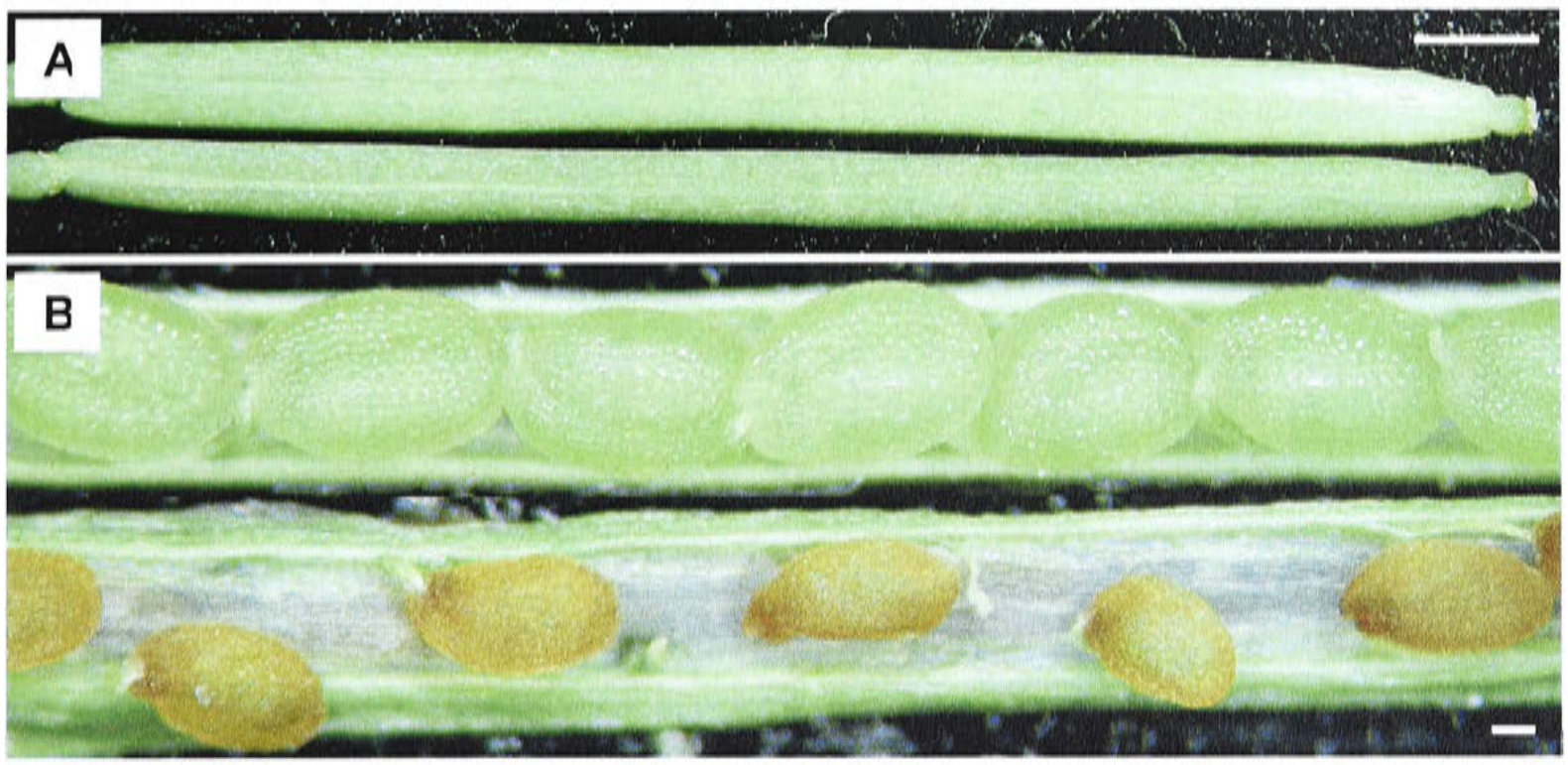


Figure 6.4 Normal silique development after crossing the MGSE-Gal4/VP16 line with the UAS-DTA line.

(**A**) Siliques from the selfed (top) and the crossed (bottom) flowers of the MGSE-Gal4/VP16 line. The siliques display no difference in both color and shape in the process of development and maturity.

(**B**) The same siliques in (**A**) show normal seed setting from the selfed silique (top) but completely abortive seeds from the crossed silique (bottom).

Bar = 1 mm in **A**; 100 μm in **B**.

6.2.4 Seed development following the reciprocal crosses between the activator line and the UAS-DTA line

The results obtained in 6.2.3 clearly show that the cross between the activator line and the UAS-DTA line lead to the production of abortive (shriveled) seeds. It is then interesting to see whether the timing of the formation of the abortive seeds in the reciprocal crosses is different. Since the reproductive development from fertilization to seed desiccation is usually completed in two weeks in *Arabidopsis*, experiments were designed to cross enough flowers so that well developed crossed siliques can be collected in one day interval for 14 days and dissected to expose seeds at different developmental stages for observation. Selfed seeds were used as control for

determining the developmental stages. Synchronization of the seed development was achieved by emasculating flowers for selfing and pollinating with pollen from the same plant at the same time when cross was carried out.

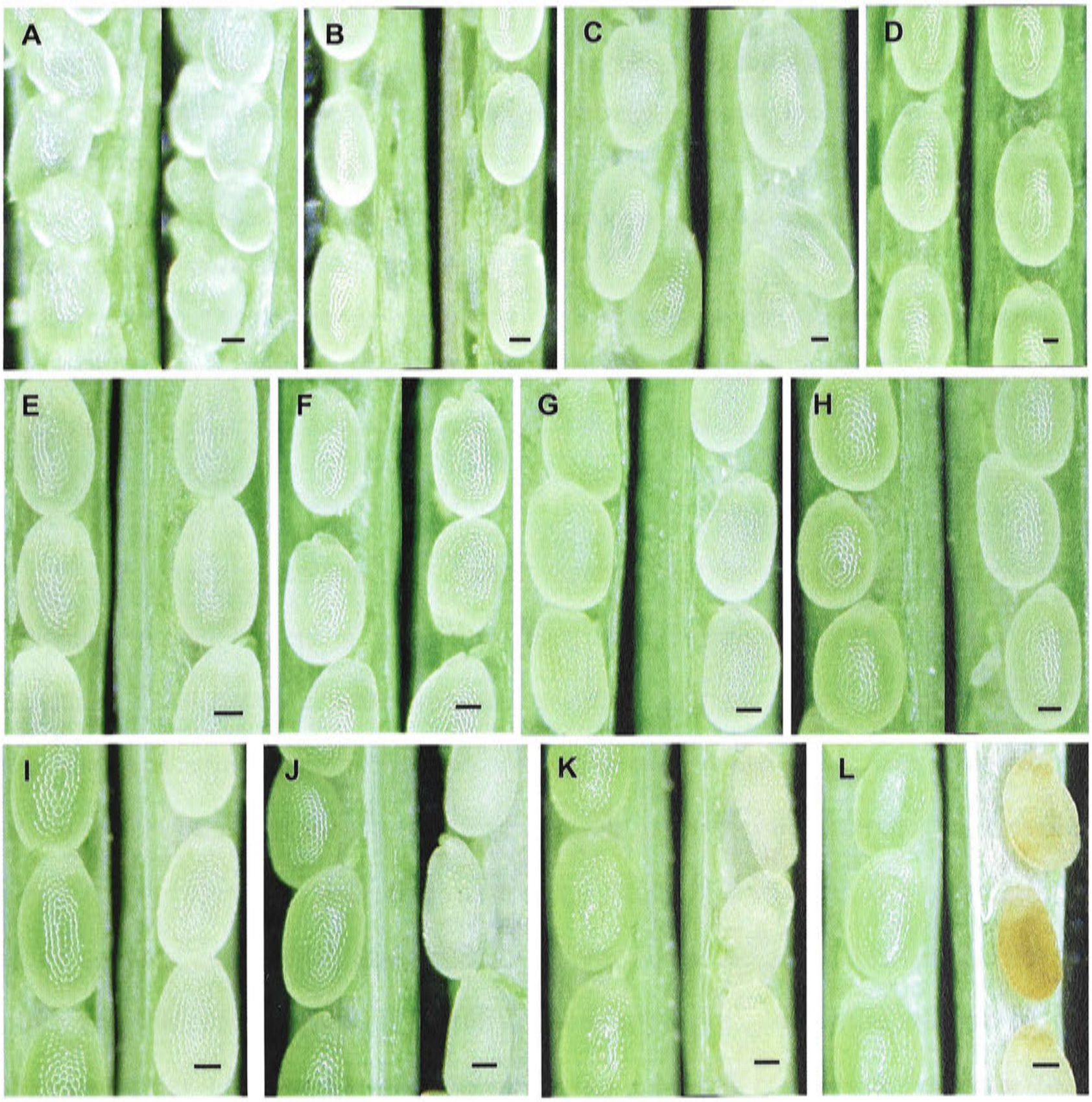


Figure 6.5 A chronological view of the seed development after pollinating the MGSE-Gal4/VP16 line (pWY-M88.2 transformed line) with pollens from the UAS-DTA line.

Seed development was recorded as days after pollination (DAP), which is defined as the numbers of days following the emasculation and pollination. To synchronize the selfed and the crossed seed development, flowers from the MGSE-Gal4/VP16 line for either self or cross were emasculated and then pollinated with pollens from the same plant (self) or pollens from the UAS-DTA

line (cross) at the same time. In each picture, the silique on the left is selfed and the silique on the right is crossed.

(**A-F**) 2, 3, 4, 5, 6 and 7 DAP seeds show no visual difference between self and cross.

(**G**) 8 DAF seeds start to show the color difference between self and cross. The crossed seeds start to display a whitish color while the selfed seeds stay green.

(**H-I**) 9 and10 DAP seeds show wider color contrast between self and cross. The selfed seeds grow greener whereas the crossed seeds become paler in the progression of development.

(**J**) 11 DAP seeds. The crossed seeds start to shrink and become flat in shape due to the lack of embryo.

(**K-L**) 12 and 14 DAP seeds. Dark pigment started to be produced in the crossed seeds prior to desiccation, while the selfed seeds remain green and filled.

Scale bar = 50 μm in **A**, **B**, **C**, **D**; 100 μm in **E**, **F**, **G**, **H**, **I**, **J**, **K**, **L**.

The events in the process of seed development were recorded as days after pollination (DAP), which is similar to days after flowering (DAF) (Bowman and Mansfield, 1994) but more accurate in terms of the starting point for fertilization. Figure 6.5 shows the development of both the crossed and the selfed seeds from 2 DAF to 14 DAF when the activator line was used as female plant. It can be seen that the crossed seeds appear normal in both color and shape up to 7 DAP when compared to the selfed seeds. From 8 DAF, obvious color difference can be seen between the selfed seeds and the crossed seeds. The selfed seeds grow greener but the crossed seeds show pale color, although the average size of the two types of seeds is the same. This trend keeps until around 11 DAF when the crossed seeds become flat in shape, indicating the lack of proper embryos in the crossed seeds. After this point, the crossed seeds start to shrivel and accumulate dark pigment (Figure 6.5 K, L), a common feature for seeds from embryo-defective mutants (Meinke, 1994).

The result from the reverse cross using the UAS-DTA line as female is quite the same (Figure 6.6). The timing of the change of color and shape of the crossed seeds

shows no clear difference from that obtained with the activator line as female. It therefore indicates that the process of the transactivation of the *DTA* gene in the reciprocal crosses is very similar.

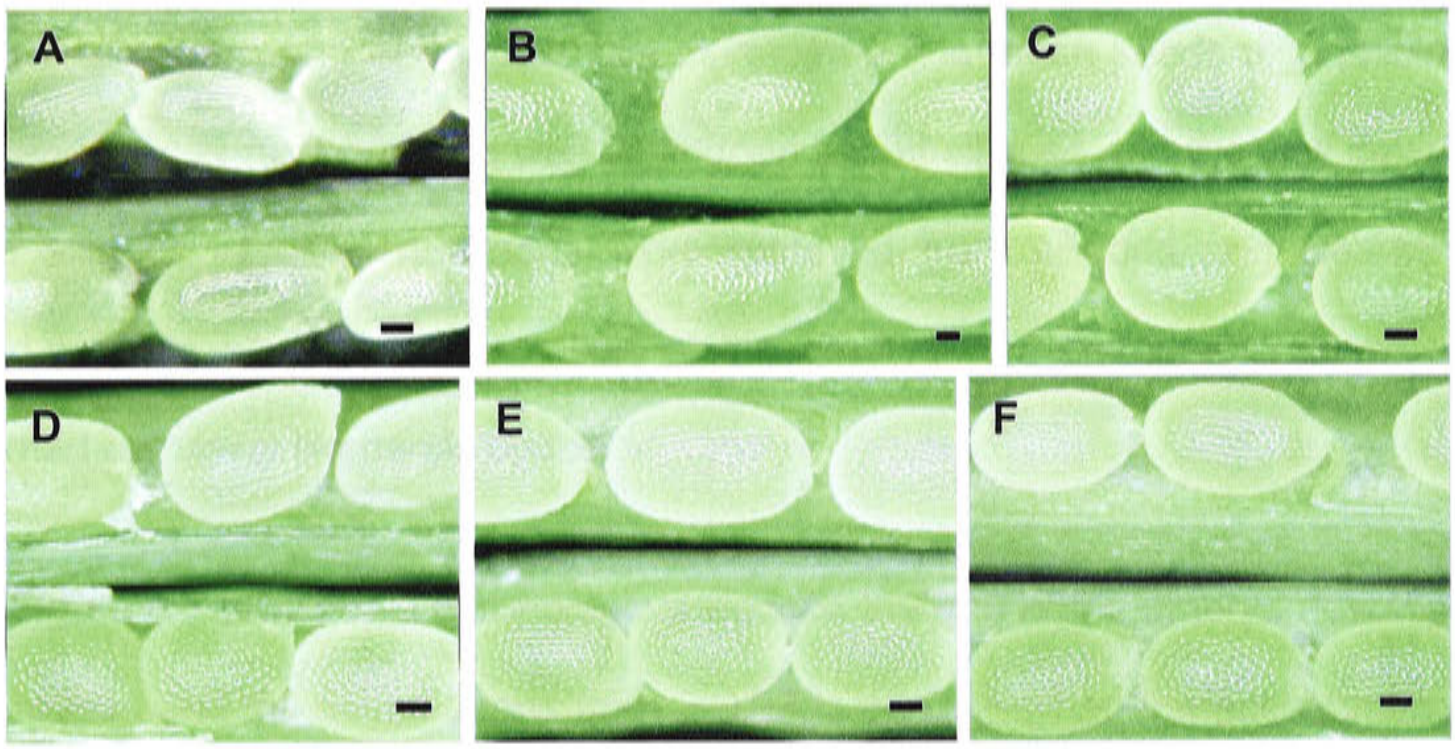


Figure 6.6 Seed development of the UAS-DT line following the pollination with pollens from the MGSE-Gal4/VP16 line.

Seed development was recorded the same as in Figure 6.5. Flowers from the UAS-DTA line for either self or cross were emasculated and then pollinated with pollens from the same plant (self) or pollens from the MGSE-Gal4/VP16 line (cross) at the same time. In each picture, the top silique is crossed and the bottom silique is selfed.

(**A-C**) 3, 5 and 7 DAF seeds show no visual difference between self and cross.

(**D**) 8 DAF seeds start to show contrast between self and cross. The selfed seeds are green whereas the crossed seeds show whitish color, the same as in Figure 6.5.

(**E**) and (**F**) 9 and10 DAP seeds show further wider color difference between self and cross. The selfed seeds grow greener but the crossed seeds become paler.

Scale bar = 50 μm in **A**, **B**; 100 μm in **C**, **D**, **E**, **F**.

6.2.5 Embryo and endosperm development following the reciprocal crosses between the activator line and the UAS-DTA line

It has been shown in 6.2.4 that the seed development seems to be normal up to 7 days after the cross between the activator line and the UAS-DTA line. To investigate whether the embryo and the endosperm development in these seeds also follows the same trend, seeds from both the well developed selfed siliques and the well developed crossed siliques at different stages were removed from siliques and cleared in a modified Hoyer's solution. Observation of the embryo and endosperm development was carried out using Nomarski optics.

Figure 6.7 provides a chronological view on the fate of the embryo development of the crossed and the selfed seeds from 2 DAF to 12 DAF when the activator line was used as female plant. Following pollination, the embryo development in the crossed seeds is normal from 1 DAF to 4 DAF that covers the zygote formation (data not shown), single- or two-terminal cell embryo (Figure6.7A), dermatogen embryo (Figure6.7C), early-globular embryo (Figure6.7E) when compared to the selfed seeds at same stages (Figure6.7B, D, F). The difference of embryo development between the crossed and the selfed seeds starts to show from 5 DAP. At this time, the embryos of the selfed seeds have advanced to early heart stage (Figure6.7H) but the embryos of the crossed seeds stay at globular stage (Figure6.7G). From 6 DAP to 12 DAP, the selfed seeds have their embryos grow from heart stage through to torpedo stage, walking-stick stage and upturned-U stage towards maturity (Figure 6.7J, L, N, P, R). In contrast, the crossed seeds go through an embryo degeneration process in the same period. In these seeds, the structure of the globular embryo is maintained up to 8 DAP (Figure 6.7I, K). Embryo degradation then becomes obvious, starting from the embryo-proper and hypophyseal cells (Figure 6.7K) to the suspensor cells (Figure 6.7M, O). This is consistent with the observation in 6.2.4 that the crossed seeds start to shrivel from 8 DAP. At 12 DAP, the embryos in the crossed seeds are not visible any more.

Meanwhile, the endosperm and the seed coat of the crossed seeds develop normally in the whole process as compared to the selfed seeds.

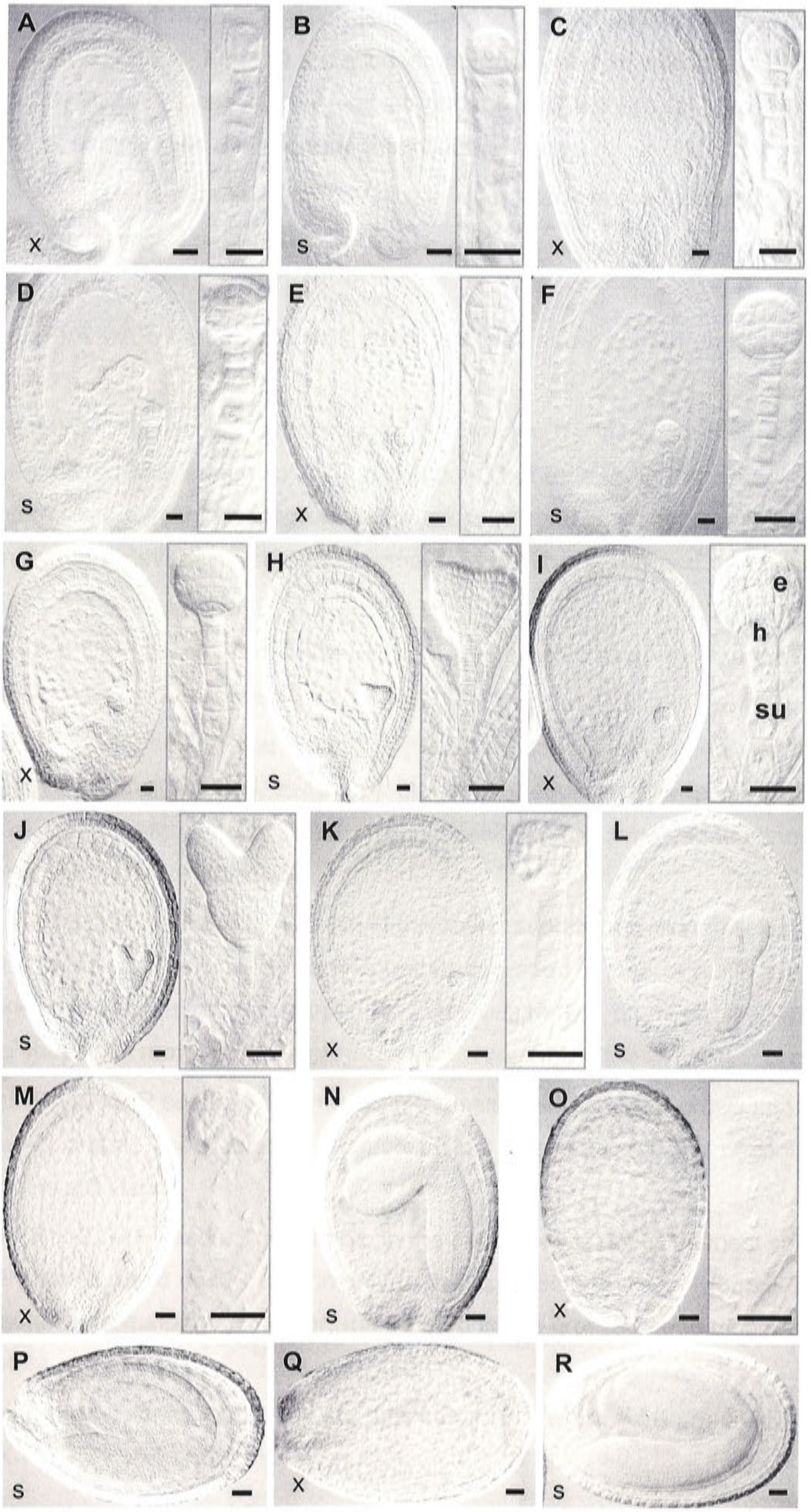


Figure 6.7 Embryo development of the MGSE-Gal4/VP16 line following the pollination with the UAS-DTA line.

Seeds at different stages were dissected from their siliques, cleared for several hours in a modified Hoyer's solution (see Chapter 2), and viewed with Nomarski optics. **A**, **C**, **E**, **G**, **I**, **K**, **M**, **O**, **Q** are seeds from cross (marked with x) and **B**, **D**, **F**, **H**, **J**, **L**, **N**, **P**, **R** are seeds from self as control (marked with s). Close-up of embryos at early stages are shown in the insets.

(**A**) and (**B**) 2 DAP seeds show normal single- or two-terminal cell embryos.

(**C**) and (**D**) 3 DAP seeds from both cross and self show normal dermatogen embryos.

(**E**) and (**F**) 4 DAP seeds from both cross and self show normal early-globular embryos.

(**G**) and (**H**) 5 DAP seeds. From this point, the embryo development of the selfed and the crossed seeds start to show difference. The embryo in the selfed seed has entered into early heart stage while the crossed seed stays at globular stage. The endosperm development is still normal for both the selfed and the crossed seeds.

(**I**) and (**J**) 6 DAP seeds. The selfed seed is at late heart stage but the crossed seed still stays at globular stage.

(**K**) and (**L**) 8 DAP seeds. The embryo in the selfed seed is at torpedo stage but the embryo in the crossed seed still stays at globular stage and the cell boundaries in the embryo-proper and hypophyseal have become vague although the suspensor cells are still intact.

(**M**) and (**N**) 9 DAP seeds. The embryo in the selfed seed is at walking-stick stage while the embryo in the crossed seed is further degraded to lose its basic cellular structure in both the embryo-proper and the suspensor although the entire embryo shape still remains.

(**O**) and (**P**) 10 DAP seeds. The embryo in the selfed seed further grows into the upturned-U stage. The degraded embryo in the crossed seed has become a bump but is still visible.

(**Q**) and (**R**) 12 DAP seeds. The embryo in the selfed seeds occupies most space inside the testa (seed coat) due primarily to the expansion of the cotyledons and is ready towards maturity. However, the embryo in the crossed seed disappears completely.

e, embryo-proper; h, hypophyseal; su, suspensor. Scale bar = 20 μm in **A**-**J**; 40 μm in **K**-**R**.

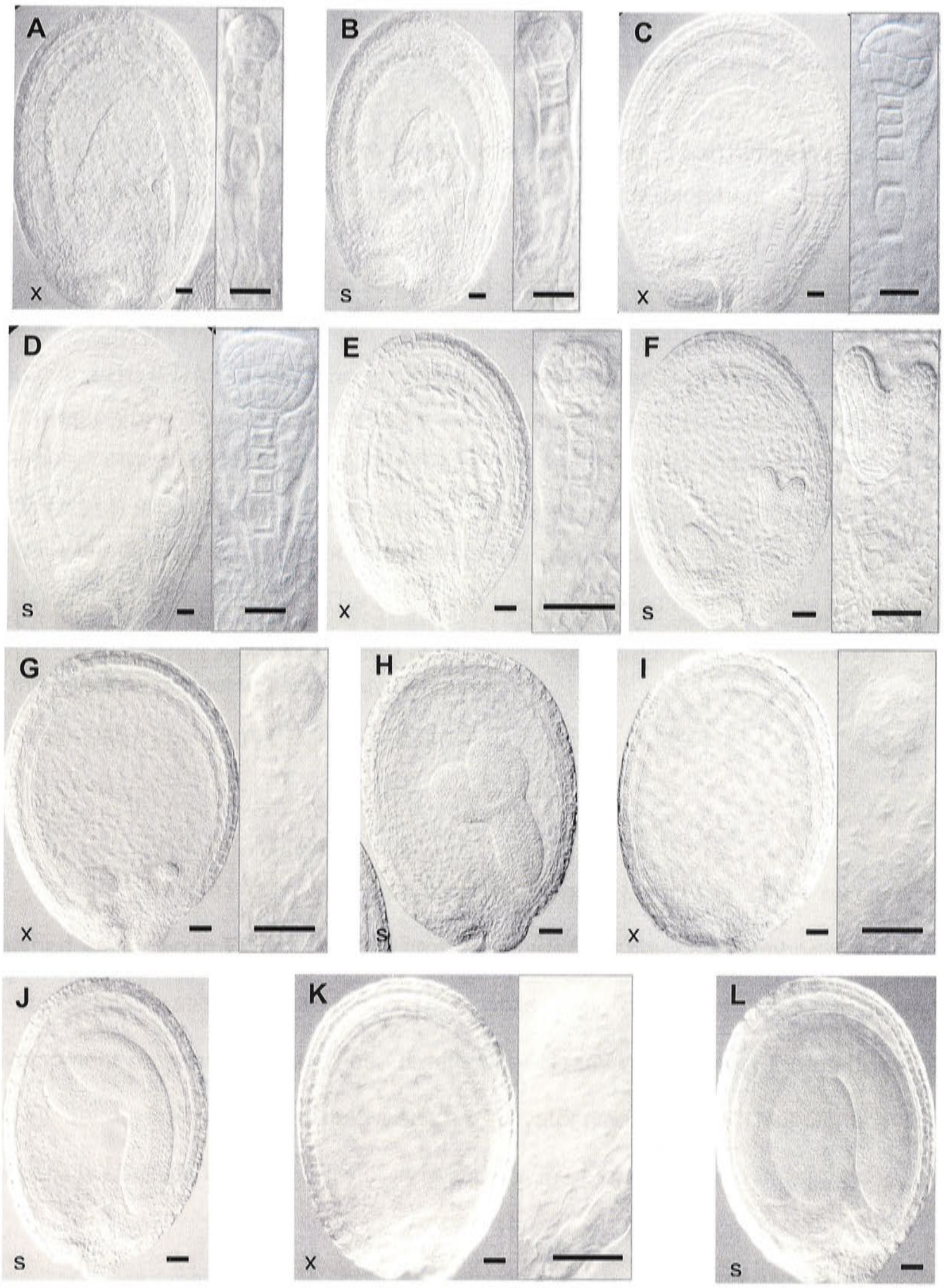


Figure 6.8 Embryo development of the UAS-DTA line following the pollination with the MGSE-Gal4/VP16 line.

Seed sample preparation and microscopy are the same as in Figure 6.6. **A**, **C**, **E**, **G**, **I**, **K** are seeds from cross (x) and **B**, **D**, **F**, **H**, **J**, **L** are seeds from self as

control (s) for comparing with seeds at the same stages in Figure 6.6. Close-up of embryos at early stages are shown in the insets.

(**A**) and (**B**) 3 DAP seeds from both cross and self show normal dermatogen embryos.

(**C**) and (**D**) 5 DAP seeds. The selfed seed is at early heart stage while the crossed seed stays at globular stage. The endosperm development is normal for both the selfed and the crossed seeds.

(**E**) and (**F**) 7 DAP seeds. The selfed seed is at late heart stage but the crossed seed still stays at globular stage.

(**G**) and (**H**) 8 DAP seeds. The embryo in the selfed seed is at torpedo stage but the embryo in the crossed seed has lost its basic cellular structure in both the embryo-proper and the suspensor although the entire embryo shape still remains.

(**I**) and (**J**) 9 DAP seeds. The embryo in the selfed seed is at walking-stick stage while the embryo in the crossed seed is further degraded.

(**K**) and (**L**) 10 DAP seeds. The embryo in the selfed seed further grows into the upturned-U stage towards maturity and the embryo in the crossed seeds stays the same before being completely degraded.

Scale bar = 20 μm in **A**, **B**, **C**, **D**; 40 μm in **E**, **F**, **G**, **H**, **I**, **J**, **K**, **L**.

When the UAS-DTA line was used as female plant, the same result was obtained as shown in Figure 6.8. The crossed seeds also have their embryos arrested at the globular stage at around 5 DAF while their endosperm and seed coat development remains normal. These results very clearly suggest that the embryos of the F1 seeds from the cross between the activator line and the UAS-DT line are specifically ablated. The specificity of the MGSE is therefore proven to be very good.

6.3 Discussion

6.3.1 The activator construct

When the transformation vector pWY-M88.2 was designed for the construction of activator lines, the 318 bp fragment and the shorter 77 bp version with the MGSE activity were the candidates for the assembling of an artificial megagametophyte-specific promoter to control the Gal4/VP16 expression. The 318 bp fragment was finally chosen. The reason not to use the 77bp fragment was that it showed activity in both orientations while the 318 bp fragment seemed to be active only in the original orientation as in the enhancer trap line ET253 (5.2.7, chapter 5). Since the T-DNA in pWY-M88.2 contains both the activator domain and the UAS sequence under the control of two separate 35S minimal promoters (Figure 6.1b), it may cause problem in directly driving the *gusA* gene expression if the dual 77bp enhancer is used. Considering that the 77 bp fragment showed even stronger activity in reverse orientation, this direct GUS expression may cover up the GUS expression transactivated by the Gal4/VP16 protein expressed under the control of the same 77 bp enhancer. If this is true, the selection of lines with GUS expression through true transactivation for subsequent crossing would be much more difficult. The use of the 318 bp fragment, however, should prevent the direct GUS expression and ensure that the GUS activity detected represents only the expression from true transcriptional activation of Gal4/VP16.

However, the pWY-M88.2 transformed lines showed quite broad GUS expression patterns as contrast to those transgenic lines obtained from constructs such as pWY-5Δ318 and pWY-5Δ77, which show very specific GUS expression in the embryo sac (Chapter 5). As described in 6.2.1.4, this type of reporter gene expression pattern may primarily be due to the activation of report gene by the Gal4 DNA-binding domain itself. The autoactivation of the reporter genes by the Gal4 DNA-binding

domain has been reported in the two-hybrid system for studying protein-protein interactions in plants (Immink and Angenent, 2002). For example, the JAB1/CSN5 activation domain binding protein can interact with the Gal4 DNA-binding domain alone and result in false positive clones (Nordgard *et al.*, 2001). It is then not unreasonable to speculate that some proteins in certain tissues in *Arabidopsis* may interact with the Gal4 DNA-binding domain, resulting in GUS expression in those tissues. For this consideration, new transcription activation system that has no background activation needs to be tested in *Arabidopsis* for better specificity.

6.3.2 Sensitivity issue for the localization of GUS or GFP expression

It has been discussed in 5.2.8 that the GUS expression in the pWY-5Δ77R transformed lines has an apparent dosage effect when lines with different copy numbers of T-DNA insertions were compared. In the case of EGFP expression as described in 5.2.9, the number of ovules with visually detectable EGFP in a single gynoecium among different pWY-O93.1 transformed lines also varies. Although the pWY-M88.2 transformed lines displayed broad GUS expression patterns as discussed in 6.3.1, the lines showing GUS expression in the embryo sac also seem to have such effect. However, according to my observation, not a single such lines show GUS or EGFP expression in all ovules in any single gynoecium or young silique even when they are homozygous. For instance, the homozygous pWY-M88.2 transformed line14 shows obvious GUS staining in only a few ovules in any best silique selected (data not shown).

In contrast, the cell ablation induced by the cross between the pWY-M88.2 transformed line14, when it was used as the activator line, and the UAS-DTA line normally leads to 100% abortive seeds in a single crossed silique (refer to 6.2.3). This means that all ovules of the activator line should have GUS expression at certain stages, but only those with strong enough expression can be detected by GUS staining. Another possibility is that the ovules in any single gynoecium are at different

developmental stages with different levels of GUS expression and only those ovules at right stage(s) can display visible GUS staining. For example, ovules at earlier stages show less intensity of GUS staining due to less tissue mass and therefore GUS expression at those stages is more difficult to detect. By this consideration, it is then possible that some of the four pWY-M88.2 transformed lines that showed no detectable GUS expression, as well as those pWY-5Δ77 transformed lines as described in Chapter 5, may have GUS expression but at an undetectable level, although gene silencing is also possible. This well supports the idea that a more sensitive reporter system may be necessary for detection of early events in female gametophyte development when new enhancer trap or gene trap system is designed as discussed in 3.3.4.

6.3.3 The MGSE is probably egg specific

Although the endogenous function of the MGSE in *Arabidopsis* genome is still unknown, the experimental results presented in Chapter 5 have shown that its capability in controlling gene expression in the megagametophyte is clear. Without conjugation with a minimal promoter, the MGSE itself is not functional (data not shown). However, when it is fused to the CaMV 35S minimal promoter, the MGSE can drive GUS or GFP reporter gene expression in a very specific manner in the embryo sac.

The cell ablation test using the DTA toxin gene under indirect control of the MGSE, however, showed that the cell ablation started in the embryo at the globular stage as demonstrated in 6.2.5. This seems to be contrary to the results from the reporter gene expression. If the transactivation of the *DTA* gene commences right after fertilization, the cell ablation should occur at the zygote formation stage. Since the test was set by hybridization between the activator line and the UAS-DTA line, the delay in cell ablation after cross could be due to delayed activation of the paternal genome during seed development as reported previously by Vielle-Calzada *et al.* (2000). In their

study, they found that, for genes expressed at later stages in seed development, the paternally inherited allele becomes active three to four days after fertilization. By looking at the embryo development in the crossed seeds described in 6.2.5, it can be seen that the DTA gene activation manner well supports this theory. In either way of crosses between the activator line and the UAS-DTA, the embryo development is arrested at globular stage around 5 days after pollination, which is close to the delayed period of the activation of paternal genome identified by Vielle-Calzada *et al.* Once the paternal genome is activated, the *DTA* gene is then transactivated, resulting in cell death in the embryo. The death of the F1 seeds is the limitation for this system, which stops the further analysis of the crossed population. An alternative system such as the glucocorticoid-mediated transcriptional induction system (Aoyama and Chua, 1997) could be tried for examining the timing of the DTA expression in developing ovules or seeds to identify the earliest time the MGSE acts.

The cell ablation results, however, clearly show that the DTA action is restricted in the embryo and the endosperm development is not affected. This strongly suggests that the MGSE direct Gal4/VP16 expression only in the egg cell before fertilization. This further fuels the idea that the MGSE could be used as a tool to study and manipulate gene expression in the egg cell for apomixis purpose.

6.3.4 Interaction of the MGSE with protein factors in the egg cell

The reciprocal crosses between the activator line and the UAS-DTA line revealed the same result that shriveled seeds were produced, suggesting that the transactivation of the *DTA* gene in both the activator line and the target line backgrounds are the same. Since the activator line has Gal4/VP16 expression in the embryo sac as revealed by the GUS staining, one can easily imagine that the activation of the *DTA* gene could take place immediately after the fertilization by pollen from the UAS-DTA line when the pollen nuclei enter the embryo sacs of the activator line.

However, if the activator line is used as male, its pollen should have no Gal4/VP16 expression in the pollen nuclei. After the activator pollen enter the embryo sacs of the UAS-DTA line, the activation of the *DTA* gene should not occur until the expression of Gal4/VP16 protein. The fact that the timing of the embryo ablation in this cross is almost the same as the cross using the activator line as female suggests the quick nature of the Gal4/VP16 activation in the embryo sac. One reasonable speculation is that certain MGSE-specific protein factor(s) are expressed in the embryo sac. When the MGSE-Gal4/VP16 DNA is delivered into the embryo sacs of the UAS-DTA line by the pollen nuclei for fertilization, the MGSE-specific protein factor(s) bind to the MGSE and lead to the expression of Gal4/VP16. The Gal4/VP16 protein can then bind to the UAS to start the expression of the *DTA* gene.

In addition, the cell ablation test suggests that the term MGSE is too broad because the MGSE has been shown to be probably egg specific as discussed in 6.3.3. To make sense, the protein factor(s) should then be expressed only in the egg cell but not in the central cell because the endosperm development would be affected by DTA expression otherwise. Therefore, a more accurate name for this enhancer might be egg-specific enhancer. More detailed work is still necessary to be done to further elucidate the nature of the MGSE.

CHAPTER 7

PROSPECTS AND FUTURE WORK

I have pursued my graduate research in three major activities: (1) isolation and characterization of a megagametophyte-specific enhancer (MGSE), (2) establishment of an enhancer testing system and (3) use of the identified enhancer to control specific gene expression in the embryo sac. When this work is put in the context of the molecular biology of the ovule and female gametophyte development in plant (Chapter I), I feel very much rewarded to have learnt and participated in the growth of this very interesting field. However, as described in Chapter I, this area of research is still in its early stages, and there is still long way to go before the molecular mechanisms underlying these processes can be completely revealed. Therefore, in this closing chapter, I would like to put forward some directions of future research, especially those closely related to the work presented in this thesis.

The first major direction of future research is the possible use of the MGSE and similar elements in apomixis. Making viable embryos without sex in seeds is a big challenge related to apomixis. The recently isolated *LEAFY COTYLEDON1 (lec1)* gene (Lotan *et al.*, 1998) has been shown to be a key gene required for the specification of cotyledon identity and the completion of embryo maturation. This gene is a central regulator of embryogenesis which can induce both the morphogenesis and maturation phases of embryogenesis but suppress vegetative development. Lotan et al. has shown that the 35S/*lec1* transgene can complement the lec1-1 mutants and lead to the production of viable seeds but seedlings from these seeds showed abnormality. The abnormalities in seedlings could be caused by the constant expression of *lec1* under the control of 35S promoter.

To make normal seeds without pollination, the MGSE could be a good candidate for controlling the *lec1* expression because it has no activity in late embryogenesis and therefore would be unlikely to affect the vegetative development. This strategy could be further tested in vaious mutants such as *fis2* (Luo *et al.*, 1999), *mea* (Grossniklaus *et al.*, 1998b) and *fie* (Ohad *et al.*, 1999) with fertilization-independent endosperm development, for viable seed formation without pollination.

A second direction of future research is to test the MGSE function in heterologous plant systems, especially in crops. *Arabidopsis thaliana* itself is only a model plant mainly for laboratory studies, but the insights from its DNA sequences and the associated functions could be very useful to crop sciences. Its crop relatives in the brassica family, such as cabbage, cauliflower and canola, shared a common ancestor with *Arabidopsis* some 12 million to 19 million years ago (Walbot, 2000). It is therefore probable that the MGSE is also present in these crop species. Homologues of the MGSE element could be isolated from these crops based on the MGSE sequence and used for the regulation of gene expression. For example, quality and yield of canola seed oil could be improved by modifying the fatty acid composition through specific expression of genes affecting seed lipid metabolism under the control of the MGSE.

A more exciting prospect is to see whether the MGSE can be functional in monocots, particularly in rice. Attempts in testing some of the 35S promoter based MGSE-containing constructs, such as pWY-F68, have been made in rice but the results obtained were complicated by the interference of the cis elements in the 35S promoter as mentioned in Chapter 5. On the other hand, constructs based on 1' promoter for plant selection, such as pWY-5Δ318, have been found to be inefficient for generating transgenic rice plants. Therefore, a better transformation vector that can provide high transformation efficiency but show no intrinsic activation of reporter gene needs to be developed before the MGSE function can be confidently tested in rice.

A third direction might be to identify the possible protein factor(s) associated with the MGSE. Regulation of transcription relies at least on two primary DNA components: promoters and enhancers. Enhancers normally consist of sequence-specific transcription factor binding sites that function distal to the transcription initial site in promoters from either an upstream or downstream position (Maniatis *et al.*, 1987; Majumder *et al.*, 1997).

At present, little is known about regulatory factors that act during megagametogenesis and the earliest stages of plant embryogenesis. The specificity of the MGSE described here is probably associated directly with some specific protein factors which are specifically expressed in the embryo sac as discussed in Chapter 6. If this is the case, some strategies can be designed to isolate the MGSE-associated protein factor(s) using the MGSE as 'bait'. One possible approach is to use the electrophoretical mobility shift assay (EMSA) (Sablowski *et al.*, 1994) to first test possible interactions between the MGSE element and any nuclear protein factor(s). Once a specific binding of regulatory protein(s) to the MGSE is confirmed, the protein(s) can then be isolated by DNA affinity purification and identified by mass spectrometry analysis. If it can be achieved, new gene(s) associated with megagametogenesis can be discovered, which will possibly open up a new research area.

A fourth area might be to explore the gain-of-function strategy based on the knowledge obtained in this thesis and somewhere else. It is well known that differential expression of genes is essential for development. There is an old and fundamental idea in the evolution of development that most evolutionary changes involve the evolution of gene regulation as opposed to changes in structural/protein coding sequences. This idea has been strengthened recently by new evidences obtained with microarray technology (Oleksiak *et al.*, 2002). The findings that some enhancer elements, such as the MGSE element and the so called cryptic or spurious enhancers with unknown functions (Pereira, 2000), are scattered in the *Arabodopsis* genome

revealed the possibilities of changes in gene regulation by genome rearrangement. This further supports the theory that changes in gene regulation rather than changes in protein sequence are a driving force in plant evolution. Using this way of thinking, novel phenotypes and traits could be created by the promotion of changes in the expression patterns of the plant's own genes (Andrzej Kilian, unpublished data).

New strategies can therefore be designed to carry out a gain-of-function screen via forced misexpression of unknown genes. One example of the strategies is the use of the Gal4-based transactivation system mentioned in Chapter 6 involving the Gal4 enhancer trap lines as activators and the random UAS insertion lines as target lines. Patterns of interest for the activator lines can be selected by reporter gene expression. By crossing the selected activators to a population of target lines, gain-of-function mutants with obvious phenotypes could be screened and used as resources for gene cloning and other plant molecular biology research.

References

Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., and Moreno, R. F. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651-1656.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

Altschul, S. F., Madden, T. L. , Schaffer, A. A., Zhang, J. , Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.

Alwine, J. C., Kemp, D. J., and Stark, G. R. (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. USA* **74**, 5350-5354.

Angenent, G. C., Franken, J., Busscher, M., van Dijken, A. , van Went, J. L., Dons, H. J., and van Tunen, A. J. (1995) A novel class of MADS box genes is involved in ovule development in petunia. *Plant Cell* **7**, 1569-1582.

Aoyama, T. and Chau, N. H. (1997) A glucocorticoid-mediated transcriptional induction system in transgenic plants. *Plant J.* **11**, 605-612.

Asamizu, E., Nakamura, Y., Sato, S., and Tabata, S. (2000) A large scale analysis of cDNA in *Arabidopsis thaliana*: generation of 12,028 non-redundant expressed sequence tags from normalized and size- selected cDNA libraries. *DNA Res.* **7**, 175-180.

Azpiroz-Leehan, R. and Feldmann, K. A. (1997) T-DNA insertion mutagenesis in *Arabidopsis*: going back and forth. *Trends Genet.* **13**, 152-156.

Babiychuk, E., Fuangthong, M., Van Montagu, M., Inze, D., and Kushnir, S. (1997) Efficient gene tagging in *Arabidopsis thaliana* using a gene trap approach. *Proc. Natl. Acad. Sci. U. S. A* **94**, 12722-12727.

Bai, X., Peirson, B. N., Dong, F., Xue, C., and Makaroff, C. A. (1999) Isolation and characterization of SYN1, a RAD21-like gene essential for meiosis in *Arabidopsis*. *Plant Cell* **11**, 417-430.

Baker, S. C., Robinson-Beers, K., Villanueva, J. M., Gaiser, J. C., and Gasser, C. S. (1997) Interactions among genes regulating ovule development in *Arabidopsis thaliana*. *Genetics* **145**, 1109-1124.

Balasubramanian, S. and Schneitz, K. (2000) NOZZLE regulates proximal-distal pattern formation, cell proliferation and early sporogenesis during ovule development in *Arabidopsis thaliana*. *Development* **127**, 4227-4238.

Bancroft, I., Bhatt, A. M., Sjodin, C., Scofield, S., Jones, J. D., and Dean, C. (1992) Development of an efficient two-element transposon tagging system in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **233**, 449-461.

Bancroft, I., Jones, J. D., and Dean, C. (1993) Heterologous transposon tagging of the *DRL1* locus in *Arabidopsis*. *Plant Cell* **5**, 631-638.

Bechtold, N., Ellis, J., and Pelletier, G. (1993) *In planta Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. *C. R. Acad. Sci. Paris/Life Sci.* **316**, 1194-1199.

Bechtold, N. and Pelletier, G. (1998) In planta *Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. *Methods Mol. Biol.* **82**, 259-266.

Bellen, H. J., D'Evelyn, D., Harvey, M., and Elledge, S. J. (1992) Isolation of temperature-sensitive diphtheria toxins in yeast and their effects on Drosophila cells. *Development* **114**, 787-796.

Bellen, H. J., O'Kane, C. J., Wilson, C., Grossniklaus, U. , Pearson, R. K., and Gehring, W. J. (1989) P-element-mediated enhancer detection: a versatile method to study development in Drosophila. *Genes Dev.* **3**, 1288-1300.

Benfey, P. N., Ren, L., and Chua, N. H. (1989) The CaMV 35S enhancer contains at least two domains which can confer different developmental and tissue-specific expression patterns. *EMBO J.* **8**, 2195-2202.

Benfey, P. N., Ren, L., and Chua, N. H. (1990a) Combinatorial and synergistic properties of CaMV 35S enhancer subdomains. *EMBO J.* **9**, 1685-1696.

Benfey, P. N., Ren, L., and Chua, N. H. (1990b) Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. *EMBO J.* **9**, 1677-1684.

Bhatt, A. M., Canales, C., and Dickinson, H. G. (2001) Plant meiosis: the means to 1N. *Trends Plant Sci.* **6**, 114-121.

Bhatt, A. M., Lister, C., Page, T., Fransz, P., Findlay, K., Jones, G. H., Dickinson, H. G., and Dean, C. (1999) The *DIF1* gene of *Arabidopsis* is required for meiotic chromosome segregation and belongs to the *REC8/RAD21* cohesin gene family. *Plant J.* **19**, 463-472.

Blackwood, E. M. and Kadonaga, J. T. (1998) Going the distance: a current view of enhancer action. *Science* **281**, 61-63.

Botella, M. A., Quesada, M. A. , Kononowicz, A. K., Bressan, R. A., Pliego, F., Hasegawa, P. M., and Valpuesta, V. (1994) Characterization and in situ localization of a salt-induced tomato peroxidase mRNA. *Plant Mol. Biol.* **25**, 105-114.

Bougourd, S., Marrison, J., and Haseloff, J. (2000) Technical advance: an aniline blue staining procedure for confocal microscopy and 3D imaging of normal and perturbed cellular phenotypes in mature *Arabidopsis* embryos. *Plant J.* **24**, 543-550.

Bouman, F (1984) The ovule. In *Embryology of Angiosperms* (Johri, B. M., ed.). Berlin: Springer-Verlag , pp. 123-158.

Bowman, J. L and Mansfield, S. G. (1994) Embryogenesis: Introduction. In *Arabidopsis: an atlas of morphology and development* (Bowman, J. L., ed.). New York: Springer-Verlag, pp. 351-361.

Bowman, J. L., Baum, S. F., Eshed, Y., Putterill, J., and Alvarez, J. (1999) Molecular genetics of gynoecium development in *Arabidopsis*. *Curr. Top. Dev. Biol.* **45**, 155-205.

Brand, A. H. and Perrimon, N. (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401-415.

Breitman, M. L., Clapoff, S., Rossant, J., Tsui, L. C., Glode, L. M., Maxwell, I. H., and Bernstein, A. (1987) Genetic ablation: targeted expression of a toxin gene causes microphthalmia in transgenic mice. *Science* **238**, 1563-1565.

Britten, R. J. (1997) Mobile elements inserted in the distant past have taken on important functions. *Gene* **205**, 177-182.

Brown, K. (2000) The human genome business today. *Scientific American* 7, 40-45.

Byzova, M. V., Franken, J., Aarts, M. G., Almeida-Engler, J., Engler, G., Mariani, C., Lookeren Campagne, M. M., and Angenent, G. C. (1999) *Arabidopsis STERILE APETALA*, a multifunctional gene regulating inflorescence, flower, and ovule development. *Genes Dev.* **13**, 1002-1014.

Campisi, L., Yang, Y., Yi, Y., Heilig, E., Herman, B., Cassista, A. J., Allen, D. W., Xiang, H., and Jack, T. (1999) Generation of enhancer trap lines in *Arabidopsis* and characterization of expression patterns in the inflorescence. *Plant J.* **17**, 699-707.

Casacuberta, E., Casacuberta, J. M., Puigdomenech, P., and Monfort, A. (1998) Presence of miniature inverted-repeat transposable elements (MITEs) in the genome of *Arabidopsis thaliana*: characterisation of the Emigrant family of elements. *Plant J.* **16**, 79-85.

Cass, D. D., Peteya, D. J., and Robertson, B. L. (1985) Megagametophyte development in *Hordeum vulgara*. 1. Early megagametogenesis and the nature of cell wall formation. *Can. J. Bot.* **63**, 2164-2171.

Chaudhury, A. M., Craig, S., Dennis, E., and Peacock, W. (1998) Ovule and embryo development, apomixis and fertilization. *Curr. Opin. Plant Biol.* **1** , 26-31.

Chaudhury, A. M., Ming, L., Miller, C., Craig, S., Dennis, E. S., and Peacock, W. J. (1997) Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A* **94**, 4223-4228.

Chen, J. L., Huisinga, K. L., Viering, M. M., Ou, S. A., Wu, C. T., and Geyer, P. K. (2002) Enhancer action in trans is permitted throughout the Drosophila genome. *Proc. Natl. Acad. Sci. U. S. A* **99**, 3723-3728.

Cheng, S., Fockler, C., Barnes, W. M., and Higuchi, R. (1994) Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci. USA* **91**, 5695-5699.

Chin, H. G., Choe, M. S., Lee, S. H., Park, S. H., Koo, J. C., Kim, N. Y., Lee, J. J., Oh, B. G., Yi, G. H., Kim, S. C., Choi, H. C., Cho, M. J., and Han, C. D. (1999) Molecular analysis of rice plants harboring an Ac/Ds transposable element-mediated gene trapping system. *Plant J.* **19**, 615 -623.

Christensen, C. A., King, E. J , Jordan, J. R., and Drews, G. N. (1997) Megagametogenesis in *Arabidopsis* wild type and the *Gf* mutant. *Sex Plant Reprod* **10**, 49-64.

Christensen, C. A., Subramanian, S., and Drews, G. N. (1998) Identification of gametophytic mutations affecting female gametophyte development in *Arabidopsis*. *Dev. Biol.* **202**, 136-151.

Clough, S. J. and Bent, A. F. (1998a) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735-743.

Clough, S. J. and Bent, A. F. (1998b) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735-743.

Colombo, L., Franken, J., Koetje, E., van Went, J., Dons, H. J., Angenent, G. C., and van Tunen, A. J. (1995) The petunia MADS box gene *FBP11* determines ovule identity. *Plant Cell* **7**, 1859-1868.

Couteau, F., Belzile, F., Horlow, C., Grandjean, O., Vezon, D., and Doutriaux, M. P. (1999) Random chromosome segregation without meiotic arrest in both male and female meiocytes of a dmc1 mutant of *Arabidopsis*. *Plant Cell* **11**, 1623-1634.

Curtis, M. D., Rae, A. L., Rusu, A. G., Harrison, S. J., and Manners, J. M. (1997) A peroxidase gene promoter induced by phytopathogens and methyl jasmonate in transgenic plants. *Mol. Plant Microbe Interact.* **10**, 326-338.

D'Halluin, K., De Block, M., Denecke, J., Janssens, J., Leemans, J., Reynaerts, A., and Botterman, J. (1992) The bar gene as selectable and screenable marker in plant engineering. *Methods Enzymol.* **216**, 415-426.

Day, C. D., Galgoci, B. F., and Irish, V. F. (1995) Genetic ablation of petal and stamen primordia to elucidate cell interactions during floral development. *Development* **121**, 2887-2895.

Dolferus, R., Jacobs, M., Peacock, W. J., and Dennis, E. S. (1994) Differential interactions of promoter elements in stress responses of the *Arabidopsis* Adh gene. *Plant Physiol* **105**, 1075-1087.

Dorsett, D. (1999) Distant liaisons: long-range enhancer-promoter interactions in Drosophila. *Curr. Opin. Genet. Dev.* **9**, 505-514.

Doutriaux, M. P., Couteau, F., Bergounioux, C., and White, C. (1998) Isolation and characterisation of the *RAD51* and *DMC1* homologs from *Arabidopsis thaliana*. *Mol. Gen. Genet.* **257**, 283-291.

Dower, W. J., Miller, J. F., and Ragsdale, C. W. (1988) High efficiency transformation of E. coli by high voltage electroporation. *Nucleic Acids Res.* **16**, 6127-6145.

Drews, G. N., Lee, D., and Christensen, C. A. (1998) Genetic analysis of female gametophyte development and function. *Plant Cell* **10**, 5-17.

Eickbush, T. H. (1992) Transposing without ends: the non-LTR retrotransposable elements. *New Biol.* **4**, 430-440.

Elliott, R. C., Betzner, A. S. , Huttner, E., Oakes, M. P., Tucker, W. Q., Gerentes, D. , Perez, P., and Smyth, D. R. (1996) *AINTEGUMENTA*, an *APETALA2* -like gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth. *Plant Cell* **8**, 155-168.

Eyal, Y., Curie, C., and McCormick, S. (1995) Pollen specificity elements reside in 30 bp of the proximal promoters of two pollen-expressed genes. *Plant Cell* **7**, 373-384.

Farkas, G., Leibovitch, B. A., and Elgin, S. C. (2000) Chromatin organization and transcriptional control of gene expression in Drosophila. *Gene* **253**, 117 -136.

Feldmann, K. A. (1991) T-DNA insertion mutagenesis in *Arabidopsis*: mutational spectrum. *Plant J.* **1**, 71-82.

Feldmann, K. A., Coury, D. A., and Christianson, M. L. (1997) Exceptional segregation of a selectable marker (KanR) in *Arabidopsis* identifies genes important for gametophytic growth and development. *Genetics* **147**, 1411-1422.

Ferrandiz, C., Gu, Q., Martienssen, R., and Yanofsky, M. F. (2000) Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. *Development* **127**, 725-734.

Finnegan, D. J. (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet.* **5**, 103-107.

Fobert, P. R., Labbe, H., Cosmopoulos, J., Gottlob-McHugh, S., Ouellet, T., Hattori, J., Sunohara, G., Iyer, V. N., and Miki, B. L. (1994) T-DNA tagging of a seed coat-specific cryptic promoter in tobacco. *Plant J.* **6**, 567-577.

Foster, E., Hattori, J., Labbe, H., Ouellet, T., Fobert, P. R., James, L. E., Iyer, V. N., and Miki, B. L. (1999) A tobacco cryptic constitutive promoter, tCUP, revealed by T-DNA tagging. *Plant Mol. Biol.* **41**, 45-55.

Gaiser, J. C., Robinson-Beers, K., and Gasser, C. S. (1995) The *Arabidopsis SUPERMAN* gene mediates asymmetric growth of the outer integument of ovules. *Plant Cell* **7**, 333-345.

Gamborg, O. L., Miller, R. A., and Ojima, K. (1968) Nutrient requirements of suspension cultures of soybean root cells. *Exp. Cell. Res.* **50**, 151-158.

Gasser, C. S., Broadhvest, J, and Hauser, B. A. (1998) Genetic analysis of ovule development. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 1- 24.

Gates, J. and Thummel, C. S. (2000) An enhancer trap screen for ecdysone-inducible genes required for Drosophila adult leg morphogenesis. *Genetics* **156**, 1765-1776.

Geyer, P. K. (1997) The role of insulator elements in defining domains of gene expression. *Curr. Opin. Genet. Dev.* **7**, 242-248.

Goff, S. A., Cone, K. C., and Fromm, M. E. (1991) Identification of functional domains in the maize transcriptional activator C1: comparison of wild-type and dominant inhibitor proteins. *Genes Dev.* **5**, 298-309.

Gossler, A., Joyner, A. L., Rossant, J., and Skarnes, W. C. (1989) Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244**, 463-465.

Greenfield, L., Bjorn, M. J., Horn, G., Fong, D., Buck, G. A., Collier, R. J., and Kaplan, D. A. (1983) Nucleotide sequence of the structural gene for diphtheria toxin carried by corynebacteriophage beta. *Proc. Natl. Acad. Sci. U. S. A* **80**, 6853-6857.

Grossniklaus, U., Koltunow, A. , and van Lookeren Campagne, M (1998a) A bright future for apomixis. *Trends Plant Sci.* **3**, 415-416.

Grossniklaus, U., Moore, J., and Gagliano, W. (1995) Analysis of *Arabidopsis* ovule development and megagametogenesis using enhancer detection. In Signaling in Plant Development. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp. 138-

Grossniklaus, U. and Schneitz, K. (1998) The molecular and genetic basis of ovule and megagametophyte development. *Semin. Cell Dev. Biol.* **9**, 227-238.

Grossniklaus, U., Vielle-Calzada, J. P., Hoeppner, M. A., and Gagliano, W. B. (1998b) Maternal control of embryogenesis by *MEDEA*, a polycomb group gene in *Arabidopsis*. *Science* **280**, 446-450.

Haig, D. (1990) New perspectives on the angiosperm female gametophyte. *Bot. Rev.* **56**, 236-274.

Hajdukiewicz, P., Svab, Z., and Maliga, P. (1994) The small, versatile pPZP family of Agrobacterium binary vectors for plant transformation. *Plant Mol. Biol.* **25**, 989-994.

Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.* **166**, 557-580.

Hauser, B. A., Villanueva, J. M., and Gasser, C. S. (1998) *Arabidopsis TSO1* regulates directional processes in cells during floral organogenesis. *Genetics* **150**, 411-423.

Higashiyama, T., Kuroiwa, H., Kawano, S., and Kuroiwa, T. (1998) Guidance in vitro of the pollen tube to the naked embryo sac of torenia fournieri. *Plant Cell* **10**, 2019-2032.

Hood, E. E., Gelvin, S. B., Melchers, S., and Hoekema, A. (1993) New Agrobacterium helper plasmids for gene transfer to plants (EHA105). *Trans Res* **2**, 208-218.

Hooykaas, P. J. and Schilperoort, R. A. (1992) Agrobacterium and plant genetic engineering. *Plant Mol. Biol.* **19**, 15-38.

Howden, R., Park, S. K., Moore, J. M., Orme, J., Grossniklaus, U., and Twell, D. (1998) Selection of T-DNA-tagged male and female gametophytic mutants by segregation distortion in *Arabidopsis*. *Genetics* **149**, 621-631.

Immink, R. G. and Angenent, G. C. (2002) Transcription factors do it together: the hows and whys of studying protein-protein interactions. *Trends Plant Sci.* **7**, 531-534.

Ito, M., Iwase, M., Kodama, H., Lavisse, P., Komamine, A., Nishihama, R., Machida, Y., and Watanabe, A. (1998) A novel cis-acting element in promoters of plant B-type cyclin genes activates M phase-specific transcription. *Plant Cell* **10**, 331-341.

Iwakawa, H., Ueno, Y., Seminarti, E., Onouchi, H., Kojima, S., Tsukaya, H., Hasebe, Y., Machida, C., and Machida, Y. (2002) The ASYMMETRIC LEAVES2 gene of *Arabidopsis thaliana*, required for formation of a symmetric flat leaf lamina, encodes a member of a novel family of proteins characterized by cysteine repeats and a leucine zipper. *Plant Cell Physiol.* In press.

Jefferson, R. A. (1987) Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Mol. Biol. Rep.* **5**, 387-405.

Jefferson, R. A. (1989) The GUS reporter gene system. *Nature* **342**, 837-838.

Jefferson, R. A. (1994) Apomixis: A social revolution for agriculture. *Biotechnology and Development Monitor* 14-16.

Jefferson, R. A., Kavanagh, T. A., and Bevan, M. W. (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901-3907.

Jofuku, K. D., den Boer, B. G. , Van Montagu, M., and Okamuro, J. K. (1994) Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. *Plant Cell* **6**, 1211-1225.

Justesen, A. F., Jespersen, H. M., and Welinder, K. G. (1998) Analysis of two incompletely spliced *Arabidopsis* cDNAs encoding novel types of peroxidase. *Biochim. Biophys. Acta* **1443**, 149-154.

Kawaoka, A., Kawamoto, T., Sekine, M., Yoshida, K., Takano, M., and Shinmyo, A. (1994) A cis-acting element and a trans-acting factor involved in the wound-induced expression of a horseradish peroxidase gene. *Plant J.* **6**, 87- 97.

Kempin, S. A., Savidge, B., and Yanofsky, M. F. (1995) Molecular basis of the cauliflower phenotype in *Arabidopsis*. *Science* **267**, 522-525.

Kertbundit, S., De Greve, H., Deboeck, F., Van Montagu, M. , and Hernalsteens, J. P. (1991) In vivo random beta-glucuronidase gene fusions in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A* **88**, 5212-5216.

Kjaersgard, I. V., Jespersen, H. M., Rasmussen, S. K., and Welinder, K. G. (1997) Sequence and RT-PCR expression analysis of two peroxidases from *Arabidopsis thaliana* belonging to a novel evolutionary branch of plant peroxidases. *Plant Mol. Biol.* **33**, 699-708.

Klimyuk, V. I. and Jones, J. D. (1997) *AtDMC1*, the *Arabidopsis* homologue of the yeast *DMC1* gene: characterization, transposon-induced allelic variation and meiosis- associated expression. *Plant J.* **11**, 1-14.

Klimyuk, V. I., Nussaume, L., Harrison, K., and Jones, J. D. (1995) Novel GUS expression patterns following transposition of an enhancer trap Ds element in *Arabidopsis*. *Mol. Gen. Genet.* **249**, 357-365.

Klucher, K. M., Chow, H., Reiser, L., and Fischer, R. L. (1996) The *AINTEGUMENTA* gene of *Arabidopsis* required for ovule and female gametophyte development is related to the floral homeotic gene *APETALA2*. *Plant Cell* **8**, 137-153.

Koch, M., Bishop, J., and Mitchell-Olds, T. (1999) Molecular systematics and evolution of *Arabidopsis* and *Arabis*. *Plant Biol.* **1**, 529-537.

Koch, M. A., Haubold, B., and Mitchell-Olds, T. (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae). *Mol. Biol. Evol.* **17**, 1483-1498.

Koltunow, A. M. (1993) Apomixis: embryo sacs and embryos formed without meiosis or fertilization in ovules. *Plant Cell* **5**, 1425-1437.

Koltunow, A. M., Bicknell, R. A, and Chaudhury, A. M. (1995) Apomixis: molecular strategies for the generation of genetically identical seeds without fertilization. *Plant Physiol* **108**, 1345-1352.

Koncz, C., Martini, N., Mayerhofer, R., Koncz-Kalman, Z., Korber, H., Redei, G. P., and Schell, J. (1989) High-frequency T-DNA-mediated gene tagging in plants. *Proc. Natl. Acad. Sci. U. S. A* **86**, 8467-8471.

Lavid, N., Schwartz, A., Yarden, O., and Tel-Or, E. (2001) The involvement of polyphenols and peroxidase activities in heavy-metal accumulation by epidermal glands of the waterlily (Nymphaeaceae). *Planta* **212**, 323-331.

Leon-Kloosterziel, K. M., Keijzer, C. J., and Koornneef, M. (1994) A seed shape mutant of *Arabidopsis* that is affected in integument development. *Plant Cell* **6**, 385-392.

Liljegren, S. J., Ditta, G. S. , Eshed, Y., Savidge, B., Bowman, J. L., and Yanofsky, M. F. (2000) *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* **404**, 766-770.

Lindsey, K., Wei, W., Clarke, M. C., McArdle, H. F., Rooke, L. M., and Topping, J. F. (1993) Tagging genomic sequences that direct transgene expression by activation of a promoter trap in plants. *Transgenic Res.* **2**, 33-47.

Liu, Y. G., Mitsukawa, N., Oosumi, T., and Whittier, R. F. (1995) Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* **8**, 457-463.

Liu, Z. and Meyerowitz, E. M. (1995) *LEUNIG* regulates *AGAMOUS* expression in *Arabidopsis* flowers. *Development* **121**, 975-991 .

Liu, Z., Running, M. P., and Meyerowitz, E. M. (1997) TSO1 functions in cell division during *Arabidopsis* flower development. *Development* **124**, 665-672.

Lloyd, A. M., Barnason, A. R., Rogers, S. G., Byrne, M. C. , Fraley, R. T., and Horsch, R. B. (1986) Transformation of *Arabidopsis thaliana* with *Agrobacterium tumefaciens*. *Science* **234**, 464-466.

Lonnig, W. E. and Saedler, H. (1997) Plant transposons: contributors to evolution? *Gene* **205**, 245-253.

Lotan, T., Ohto, M., Yee, K. M., West, M. A., Lo, R., Kwong, R. W., Yamagishi, K., Fischer, R. L., Goldberg, R. B., and Harada, J. J. (1998) *Arabidopsis LEAFY COTYLEDON1* is sufficient to induce embryo development in vegetative cells. *Cell* **93**, 1195-1205.

Lukacsovich, T., Asztalos, Z., Awano, W., Baba, K., Kondo, S., Niwa, S., and Yamamoto, D. (2001) Dual-tagging gene trap of novel genes in Drosophila melanogaster. *Genetics* **157**, 727-742.

Luo, M., Bilodeau, P., Koltunow, A., Dennis, E. S., Peacock, W. J., and Chaudhury, A. M. (1999) Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A* **96**, 296-301.

Lyko, F. and Paro, R. (1999) Chromosomal elements conferring epigenetic inheritance. *Bioessays* **21**, 824-832.

Ma, H. (1997) The on and off of floral regulatory genes. *Cell* **89**, 821-824.

Ma, J., Przibilla, E., Hu, J., Bogorad, L., and Ptashne, M. (1988) Yeast activators stimulate plant gene expression. *Nature* **334**, 631-633.

Maheshwari, P (1950) *An Introduction to the Embryology of Angiosperms*. New York: McGraw-Hill.

Mahmoudi, T., Katsani, K. R., and Verrijzer, C. P. (2002) GAGA can mediate enhancer function in trans by linking two separate DNA molecules. *EMBO J.* **21**, 1775-1781.

Majumder, S., Zhao, Z., Kaneko, K., and DePamphilis, M. L. (1997) Developmental acquisition of enhancer function requires a unique coactivator activity. *EMBO J.* **16**, 1721-1731.

Maniatis, T., Goodbourn, S., and Fischer, J. A. (1987) Regulation of inducible and tissue-specific gene expression. *Science* **236**, 1237-1245.

Mansfield, S. G. and Bowman, J. L. (1994) Ovules: Introduction. In *Arabidopsis: an atlas of morphology and development* (Bowman, J. L., ed.). New York: Springer-Verlag, pp. 299-301.

Mansfield, S. G., Briarty, L. G., and Erni, S. (1991) Early embryogenesis in *Arabidopsis thaliana*. I. The mature embryo sac. *Can. J. Bot.* **69**, 447-460.

Martienssen, R. A. (1998) Functional genomics: probing plant gene function and expression with transposons. *Proc. Natl. Acad. Sci. U. S. A* **95**, 2021-2026.

Mayer, K., Schuller, C., Wambutt, R., Murphy, G., Volckaert, G., Pohl, T., Dusterhoft, A., Stiekema, W., Entian, K. D., Terryn, N., Harris, B., Ansorge, W., Brandt, P., Grivell, L., Rieger, M., Weichselgartner, M., De, Simone, V, Obermaier, B., Mache, R., Muller, M., Kreis, M., Delseny, M., Puigdomenech, P., Watson, M., McCombie, W. R., and . (1999) Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* **402**, 769-777.

Mayer, U. and Jurgens, G. (1998) Pattern formation in plant embryogenesis: a reassessment. *Semin. Cell Dev. Biol.* **9**, 187-193.

McClintock, B. (1950) The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* **6**, 344-355.

Meinke, D. W. (1994) Seed development in *Arabidopsis thaliana*. In *Arabidopsis* (Meyerowitz, E. M. and Somerville, C. R., ed.). Cold Spring Harbor: Cold Spring Harbor Laboratory Press, pp. 253-295.

Mengiste, T., Amedeo, P., and Paszkowski, J. (1997) High-efficiency transformation of *Arabidopsis thaliana* with a selectable marker gene regulated by the T-DNA 1' promoter. *Plant J.* **12**, 945-948.

Mercier, R., Vezon, D., Bullier, E., Motamayor, J. C., Sellier, A., Lefevre, F., Pelletier, G., and Horlow, C. (2001) SWITCH1 (SWI1): a novel protein required for the establishment of sister chromatid cohesion and for bivalent formation at meiosis. *Genes Dev.* **15**, 1859-1871.

Misra, R. C. (1962) Contribution to the embryology of *Arabidopsis thaliana* (Gay and Monn.). *Agra Univ. J. Res. Sci.* **11**, 191-199.

Mohan, R., Bajar, A. M., and Kolattukudy, P. E. (1993) Induction of a tomato anionic peroxidase gene (*tap1*) by wounding in transgenic tobacco and activation of *tap1*/GUS and *tap2*/GUS chimeric gene fusions in transgenic tobacco by wounding and pathogen attack. *Plant Mol. Biol.* **21**, 341-354.

Mollereau, B., Wernet, M. F., Beaufils, P., Killian, D., Pichaud, F., Kuhnlein, R., and Desplan, C. (2000) A green fluorescent protein enhancer trap screen in *Drosophila* photoreceptor cells. *Mech. Dev.* **93**, 151-160.

Moore, I., Galweiler, L., Grosskopf, D., Schell, J., and Palme, K. (1998) A transcription activation system for regulated gene expression in transgenic plants. *Proc. Natl. Acad. Sci. U. S. A* **95**, 376-381.

Moore, J. M., Calzada, J. P., Gagliano, W., and Grossniklaus, U. (1997) Genetic characterization of *hadad*, a mutant disrupting female gametogenesis in *Arabidopsis thaliana*. *Cold Spring Harb. Symp. Quant. Biol.* **62**, 35-47.

Motamayor, J. C., Vezon, D., Bajon, C., Sauvanet, A., Grandjean, O., Marchand, M., Bechtold, N., Pelletier, G., and Horlow, C. (2000) *Switch* (*swi1*), an *Arabidopsis thaliana* mutant affected in the female meiotic switch. *Sex Plant Reprod* **12**, 209-218.

Murashige, T. and Skoog, F. (1962) A revised medium for rapid growth and bioassay with tobacco tissue cultures. *Physiol. Plant* **15**, 473-497.

Murray, M. G. and Thompson, W. F. (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* **8**, 4321-4325.

Nilsson, O., Wu, E., Wolfe, D. S., and Weigel, D. (1998) Genetic ablation of flowers in transgenic *Arabidopsis*. *Plant J.* **15**, 799-804.

Nogler, G. A. (1984) Gametophytic apomixis. In *Embryology of Angiosperms* (Johri, B. M., ed.). Berlin: Springer-Verlag, pp. 475-518.

Nordgard, O., Dahle, O., Andersen, T. O., and Gabrielsen, O. S. (2001) JAB1/CSN5 interacts with the GAL4 DNA binding domain: a note of caution about two-hybrid interactions. *Biochimie* **83**, 969-971.

O'Kane, C. J. and Gehring, W. J. (1987) Detection in situ of genomic regulatory elements in Drosophila. *Proc. Natl. Acad. Sci. U. S. A* **84**, 9123-9127.

Ochman, H., Ayala, F. J., and Hartl, D. L. (1993) Use of polymerase chain reaction to amplify segments outside boundaries of known sequences. *Methods Enzymol.* **218**, 309-321.

Ochman, H., Gerber, A. S., and Hartl, D. L. (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621-623.

Ochman, H., Medhora, M. M., Garza, D, and Hartl, D. L. (1990) Amplification of flanking sequences by inverse PCR. In *PCR Protocols: A Guide to Methods and Applications* (Innis, M. A., ed.). San Diago: Academic Press, pp. 219-227.

Ohad, N., Margossian, L., Hsu, Y., Williams, C., Repetti, P., and Fischer, R. L. (1996) A Mutation that Allows Endosperm Development without Fertilization. *Proc. Natl. Acad. Sci. U. S. A* **93**, 5319-5324.

Ohad, N., Yadegari, R., Margossian, L., Hannon, M., Michaeli, D., Harada, J. J., Goldberg, R. B., and Fischer, R. L. (1999) Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. *Plant Cell* **11**, 407-416.

Oleksiak, M. F., Churchill, G. A., and Crawford, D. L. (2002) Variation in gene expression within and among natural populations. *Nat. Genet.* **32**, 261-266.

Omann, F., Beaulieu, N., and Tyson, H. (1994) cDNA sequence and tissue-specific expression of an anionic flax peroxidase. *Genome* **37**, 137-147.

Osborne, B. I. and Baker, B. (1995) Movers and shakers: maize transposons as tools for analyzing other plant genomes. *Curr. Opin. Cell Biol.* **7**, 406-413.

Ostergaard, L., Pedersen, A. G., Jespersen, H. M., Brunak, S., and Welinder, K. G. (1998) Computational analyses and annotations of the *Arabidopsis* peroxidase gene family. *FEBS Lett.* **433**, 98-102.

Palmiter, R. D., Behringer, R. R., Quaife, C. J., Maxwell, F., Maxwell, I. H., and Brinster, R. L. (1987) Cell lineage ablation in transgenic mice by cell-specific expression of a toxin gene. *Cell* **50**, 435-443.

Parinov, S., Sevugan, M., De, Ye, Yang, W. C., Kumaran, M., and Sundaresan, V. (1999) Analysis of flanking sequences from dissociation insertion lines: a database for reverse genetics in *Arabidopsis*. *Plant Cell* **11**, 2263- 2270.

Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., and Yanofsky, M. F. (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* **405**, 200-203.

Pereira, A. (2000) A transgenic perspective on plant functional genomics. *Transgenic Res* **9**, 245-260.

Pirrotta, V. (1997) Chromatin-silencing mechanisms in Drosophila maintain patterns of gene expression. *Trends Genet.* **13**, 314-318.

pohlman, R. F., Fedoroff, N. V., and Messing, J. (1984) The nucleotide sequence of the maize controlling element *Activator*. *Cell* **37**, 635-643.

Price, R. A. and Palmer, J. D. (1994) Systematic relationships of *Arabidopsis* : a molecular and morphological perspective. In *Arabidopsis* (Meyerowitz, E. M. and Somerville, C. R., ed.). Cold Spring Harbor: Cold Spring Harbor Laboratory Press, pp. 7-19.

Rasmussen, S. K., Welinder, K. G., and Hejgaard, J. (1991) cDNA cloning, characterization and expression of an endosperm-specific barley peroxidase. *Plant Mol. Biol.* **16**, 317-327.

Ray, A., Lang, J. D., Golden, T., and Ray, S. (1996) *SHORT INTEGUMENT* (*SIN1*), a gene required for ovule development in *Arabidopsis*, also controls flowering time. *Development* **122**, 2631-2638.

Ray, A., Robinson-Beers, K., Ray, S., Baker, S. C., Lang, J. D., Preuss, D., Milligan, S. B., and Gasser, C. S. (1994) *Arabidopsis* floral homeotic gene *BELL (BEL1)* controls ovule development through negative regulation of *AGAMOUS* gene (*AG*) [see comments]. *Proc. Natl. Acad. Sci. U. S. A* **91**, 5761-5765.

Redei, G. P. (1965) Non-mendelian megagametogenesis in *Arabidopsis*. *Genetics* **51**, 857-872.

Reiser, L. and Fischer, R. L. (1993) The ovule and the embryo sac. *Plant Cell* **5**, 1291-1301.

Reiser, L., Modrusan, Z., Margossian, L., Samach, A., Ohad, N., Haughn, G. W., and Fischer, R. L. (1995) The *BELL1* gene encodes a homeodomain protein involved in pattern formation in the *Arabidopsis* ovule primordium. *Cell* **83**, 735-742.

Rio, A., Puigdomenech, P., and Casacuberta, J. M. (1996) Mrs, a new subfamily of Tourist transposable elements. *Plant Mol. Biol.* **32**, 1221-1226.

Robinson-Beers, K., Pruitt, R. E., and Gasser, C. S. (1992) Ovule development in wild-type *Arabidopsis* and two female-sterile mutants. *Plant Cell* **4**, 1237-1249.

Roe, J. L., Nemhauser, J. L., and Zambryski, P. C. (1997) *TOUSLED* participates in apical tissue formation during gynoecium development in *Arabidopsis*. *Plant Cell* **9**, 335-353.

Russell, S. D. (1993) The egg cell: development and role in fertilization and early embryogenesis. *Plant Cell* **5**, 1349-1359.

Sablowski, R. W., Moyano, E., Culianez-Macia, F. A., Schuch, W., Martin, C., and Bevan, M. (1994) A flower-specific Myb protein activates transcription of phenylpropanoid biosynthetic genes. *EMBO J.* **13**, 128-137.

Sadowski, I., Ma, J., Triezenberg, S., and Ptashne, M. (1988) GAL4-VP16 is an unusually potent transcriptional activator. *Nature* **335**, 563-564.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A* **74**, 5463-5467.

Savidan, Y. (2000) Apomixis: genetics and breeding. *Plant Breeding Reviews* **18**, 13-82.

Scheer, N. and Camnos-Ortega, J. A. (1999) Use of the Gal4-UAS technique for targeted gene expression in the zebrafish. *Mech. Dev.* **80**, 153-158.

Schlappi, M., Raina, R., and Fedoroff, N. (1996) A highly sensitive plant hybrid protein assay system based on the Spm promoter and TnpA protein for detection and analysis of transcription activation domains. *Plant Mol. Biol.* 717-725.

Schneitz, K. (1999) The molecular and genetic control of ovule development. *Curr. Opin. Plant Biol.* **2**, 13-17.

Schneitz, K., Balasubramanian, S., and Schiefthaler, U. (1998) Organogenesis in plants: the molecular and genetic control of ovule development. *Trends Plant Sci.* **3**, 468-472.

Schneitz, K., Hulskamp, M., Kopczak, S. D., and Pruitt, R. E. (1997) Dissection of sexual organ ontogenesis: a genetic analysis of ovule development in *Arabidopsis thaliana*. *Development* **124**, 1367-1376.

Schneitz, K., Hulskamp, M., and Pruitt, R. E. (1995) Wild-type ovule development in *Arabidopsis thaliana*: a light microscope study of cleared whole-mount tissue. *Plant J.* **7**, 731-749.

Scott, L., LaFoe, D., and Weil, C. F. (1996) Adjacent sequences influence DNA repair accompanying transposon excision in maize. *Genetics* **142**, 237-246.

Sessions, A., Yanofsky, M. F., and Weigel, D. (1998) Patterning the floral meristem. *Semin. Cell Dev. Biol.* **9**, 221-226.

Sheen, J., Hwang, S., Niwa, Y., Kobayashi, H., and Galbraith, D. W. (1995) Green-fluorescent protein as a new vital marker in plant cells. *Plant J.* **8**, 777-784.

Sheridan, W. F., Avalkina, N. A., Shamrov, I. I., Batygina, T. B., and Golubovskaya, I. N. (1996) The *mac1* gene: controlling the commitment to the meiotic pathway in maize. *Genetics* **142**, 1009-1020.

Siddiqi, I., Ganesh, G., Grossniklaus, U., and Subbiah, V. (2000) The *dyad* gene is required for progression through female meiosis in *Arabidopsis*. *Development* **127**, 197-207.

Skarnes, W. C. (1990) Entrapment vectors: a new tool for mammalian genetics. *Biotechnology (N. Y.)* **8**, 827-831.

Skinner, D. J., Baker, S. C., Meister, R. J., Broadhvest, J., Schneitz, K., and Gasser, C. S. (2001) The *Arabidopsis HUELLENLOS* Gene, Which Is Essential for Normal Ovule Development, Encodes a Mitochondrial Ribosomal Protein. *Plant Cell* **13**, 2719-2730.

Somerville, C. R. and Meyerowitz, E. M. (1994) Introduction. In *Arabidopsis* (Meyerowitz, E. M. and Somerville, C. R., ed.). Cold Spring Harbor: Cold Spring Harbor Laboratory Press, pp. 1-6.

Song, J. Y., Leung, T., Ehler, L. K., Wang, C., and Liu, Z. (2000) Regulation of meristem organization and cell division by TSO1, an *Arabidopsis* gene with cysteine-rich repeats. *Development* **127**, 2207-2217.

Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517.

Springer, P. S. (2000) Gene traps: tools for plant development and genomics. *Plant Cell* **12**, 1007-1020.

Springer, P. S., McCombie, W. R., Sundaresan, V., and Martienssen, R. A. (1995) Gene trap tagging of PROLIFERA, an essential MCM2-3-5-like gene in *Arabidopsis*. *Science* **268**, 877-880.

Stanford, W. L., Cohn, J. B., and Cordes, S. P. (2001) Gene-trap mutagenesis: past, present and beyond. *Nat. Rev. Genet.* **2**, 756-768.

Stewart, A. C., Gravitt, P. E. , Cheng, S., and Wheeler, C. M. (1995) Generation of entire human papillomavirus genomes by long PCR: frequency of errors produced during amplification. *Genome Res.* **5**, 79-88.

Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J. D., Dean, C., Ma, H., and Martienssen, R. (1995) Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Genes Dev.* **9**, 1797-1810.

Surzycki, S. A. and Belknap, W. R. (1999) Characterization of repetitive DNA elements in *Arabidopsis*. *J. Mol. Evol.* **48**, 684-691.

The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815.

Thorsness, M. K., Kandasamy, M. K, Nasrallah, M. E., and Nasrallah, J. B. (1993) Genetic ablation of floral cells in *Arabidopsis*. *Plant Cell* **5**, 253-261.

Topping, J. F. and Lindsey, K. (1995) Insertional mutagenesis and promoter trapping in plants for the isolation of genes and the study of development. *Transgenic Res* **4**, 291-305.

Triezenberg, S. J., Kingsbury, R. C., and McKnight, S. L. (1988) Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev.* **2**, 718-729.

Tsugeki, R. and Fedoroff, N. V. (1999) Genetic ablation of root cap cells in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **96**, 12941-12946.

Tu, Z. (1997) Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, Aedes aegypti. *Proc. Natl. Acad. Sci. U. S. A* **94**, 7475-7480.

Valvekens, D., Montague, M. V. , and Lijsbettens, M. V. (1988) *Agrobaterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. *Proc. Natl. Acad. Sci. USA* **85**, 5536-5540.

van der Geest, A. H. M., Frisch, D. A., Kemp, J. D., and Hall, T. C. (1995) Cell ablation reveals that expression from the phaseolin promoter is confined to embryogenesis and microsporogenesis. *Plant Physiol* **109**, 1151-1158.

van, Dijk, I and van, Damme, I (2000) Apomixis technology and the paradox of sex. *Trends Plant Sci.* **5**, 81-84.

Velten, J., Velten, L., Hain, R., and Schell, J. (1984) Isolation of a dual plant promoter fragment from the Ti plasmid of *Agrobacterium tumefaciens*. *EMBO J.* **3**, 2723-2730 .

Vielle-Calzada, J. P., Baskar, R., and Grossniklaus, U. (2000) Delayed activation of the paternal genome during seed development. *Nature* **404**, 91-94.

Vielle-Calzada, J. P., Crane, C. F., and Stelly, D. M. (1996) Apomixis: the asexual revolution. *Science* **274**, 1322-1323.

Walbot, V. (2000) *Arabidopsis thaliana* genome. A green chapter in the book of life. *Nature* **408**, 794-795.

Wanapu, C. and Shinmyo, A. (1996) Cis-regulatory elements of the peroxidase gene in *Arabidopsis thaliana* involved in root-specific expression and responsiveness to high-salt stress. *Ann. N. Y. Acad. Sci.* **782**, 107-114.

Webb, M. C. and Bowman, J. L. (1994) Stages of embryo sac and ovule development. In *Arabidopsis: an atlas of morphology and development* (Bowman, J. L., ed.). New York: Springer-Verlag, pp. 300-

Webb, M. C. and Gunning, B. E. S. (1990) Embryo sac development in *Arabidopsis thaliana*. I. Megasporogenesis, including the microtubular cytoskeleton. *Sex Plant Reprod* **3**, 244-256.

Webb, M. C. and Gunning, B. E. S. (1994) Embryo sac development in *Arabidopsis thaliana*. II. The cytoskeleton during megagametogenesis. *Sex Plant Reprod* **7**, 153-163.

Wei, W. and Brennan, M. D. (2000) Polarity of transcriptional enhancement revealed by an insulator element. *Proc. Natl. Acad. Sci. U. S. A* **97**, 14518-14523.

Weigel, D. (1995) The APETALA2 domain is related to a novel type of DNA binding domain. *Plant Cell* **7**, 388-389.

Wessler, S. R. (1996) Turned on by stress. Plant retrotransposons. *Curr. Biol.* **6**, 959-961.

Wessler, S. R., Bureau, T. E., and White, S. E. (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr. Opin. Genet. Dev.* **5**, 814-821.

White, J., Chang, S. Y., Bibb, M. J., and Bibb, M. J. (1990) A cassette containing the bar gene of Streptomyces hygroscopicus: a selectable marker for plant transformation. *Nucleic Acids Res* **18**, 1062-

Willemse, M. T. M. and van Went, J. L. (1984) The female gametophyte. In *Embryology of Angiosperms* (Johri, B. M., ed.). Berlin: Springer-Verlag, pp. 159-196.

Wilson, C., Pearson, R. K., Bellen, H. J., O'Kane, C. J., Grossniklaus, U., and Gehring, W. J. (1989) P-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in Drosophila. *Genes Dev.* **3**, 1301-1313.

Xie, M., He, Y., and Gan, S. (2001) Bidirectionalization of polar promoters in plants. *Nat. Biotechnol.* **19**, 677-679.

Yang, T. T., Cheng, L., and Kain, S. R. (1996) Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. *Nucleic Acids Res.* **24**, 4592-4593.

Yang, W. C. and Sundaresan, V. (2000) Genetics of gametophyte biogenesis in *Arabidopsis*. *Curr. Opin. Plant Biol.* **3**, 53-57.

Yang, W. C., Ye, D., Xu, J., and Sundaresan, V. (1999) The *SPOROCYTELESS* gene of *Arabidopsis* is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes Dev.* **13**, 2108-2117.

Ye, G. N., Stone, D., Pang, S. Z., Creely, W., Gonzalez, K., and Hinchee, M. (1999) *Arabidopsis* ovule is the target for Agrobacterium in planta vacuum infiltration transformation. *Plant J.* **19**, 249-257.

Yephremov, A. and Saedler, H. (2000) Technical advance: display and isolation of transposon-flanking sequences starting from genomic DNA or RNA. *Plant J.* **21**, 495-505.

Zhan, H. C., Liu, D. P., and Liang, C. C. (2001) Insulator: from chromatin domain boundary to gene regulation. *Hum. Genet.* **109**, 471-478.

Zheng, K., Huang, N., Bennett, J., and Khush, G. S. (1995) *PCR-based marker-assisted selection in rice breeding.* IRRI Discussion Paper Series No.12. Manila: International Rice Research Institute

Zupan, J., Muth, T. R., Draper, O., and Zambryski, P. (2000) The transfer of DNA from agrobacterium tumefaciens into plants: a feast of fundamental insights. *Plant J.* **23**, 11-28.

APPENDIX A

List of Abbreviations Used

α	alpha
β	beta
µg	microgram
µl	microliter(s)
µM	micromolar
A	adenine
Ac/Ds	activator/dissociation transposons
ATP	adenosine-5'-triphosphate
bar	phosphinothricin acetyl transferase gene
BSA	bovine serum albumin
C	cytidine
CaMV	cauliflower mosaic virus
CTAB	cetyltrimethylammonium bromide
dATP	deoxyadenosine-5'-triphosphate
DEPC	diethyl pyrocarbonate
DNA	deoxyribonucleic acid
dNTPs	mixture of dATP, dTTP, dCTP and dGTP
DsE	enhancer trap Ds element
EDTA	ethylenediaminetetraactic acid
EGFP	enhanced green fluorescence protein
EMS	ethyl methanesulfonate
EMSA	electrophoretical mobility shift assay
g	gram
G	guanosine
GUS	β-glucuronidase
gusA	β-glucuronidase gene
hyg	hygromycin phosphotransferase gene
L	liter(s)
MCS	multiple cloning sites
MES	2-(N-morpholino)ethanesulfonic acid
MGSE	megagametophyte-specific enhancer
min	minute(s)
MITE	miniature inverted-repeat transposable element
ml	milliliter(s)

mM	millimolar
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger ribonuleic acid
ng	nanogram
nptII	neomycin phosphotransferase II gene
PIPES	piperazine-N',N'-bis(2-ethaesulfonic acid)
RNA	ribonuleic acid
RNase	ribonulease
rpm	rotation per minute
RT	room temperature
SDS	sodium dodecyl sulphate
sec	second(s)
T	thymine
T-DNA	transfer DNA
Tm	melting temperature
UAS	upstream activation sequence
UPE	upstream promoter element
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
X-Gluc	5-bromo-4-chloro-3-indolyl-β-D-glucuronide

Appendix B

The 11575 bp DsE flanking region on Chromosome 4 from the enhancer trap line ET 253 (in Landsberg erecta background)

```
FEATURES             Location/Qualifiers
     insertion_seq   8580..8587
                     /note="Site of Ds Enhancer Trap insertion"
                     /label=DsE\Insertion\site
     transposon      8585..8745
                     /label=MITE-like\transposable\element
     misc_feature    8511..8587
                     /label=Region\with\MGSE\activity
     CDS             complement(10022..10567)
                     /label=ASL41\gene\2rd\exon
     CDS             complement(10706..10888)
                     /label=ASL41\1st\exon
     CDS             2057..2272
                     /note="real ATP9a gene"
                     /label=atpH21-2\gene\1st\exon
     CDS             2355..2555
                     /note="real ATP9a gene"
                     /label=atpH21-2\gene\2nd\exon
     CDS             2629..2794
                     /note="real ATP9a gene"
                     /label=atpH21-2\gene\3rd\exon
     CDS             3007..3410
                     /note="real ATP9a gene"
                     /label=atpH21-2\gene\4th\exon
     CDS             5140..5355
                     /note="peroxidase gene"
                     /label=atpH21-1\gene\1st\exon
     CDS             5444..5644
                     /note="peroxidase gene"
                     /label=atpH21-1\gene\2nd\exon
     CDS             5734..5899
                     /note="peroxidase gene"
                     /label=atpH21-1\gene\3rd\exon
     CDS             6048..6451
                     /note=" peroxidase gene"
                     /label=atpH21-1\gene\4th\exon
ORIGIN
    1 gaattcaaaa cactattata ttatattatt ataacatatt ggaaagggtt gacaaaagaa
   61 aaacataccg gaaatgagat cctctctttc taagtaatta ttggcgttat aactttcatc
  121 ataaatcata agcacggctt ataggtatac atgtcactga aacttgtatc aaaacatgta
  181 actaacaaat atcatatgtt attgtatata tatatacaaa aatgctatga caggaagcaa
  241 tctaaagact atactcatta acgaattagg ttctacaatt acactcgatc acataaattt
  301 ctgttttata tttattacaa cccatgtatg tgtagagtct atagtttaat agtgtttaat
  361 taaaatttat cacaaaccca tgtatgtgta gagtttaata gagaaggaaa attttagtga
  421 atcgctaagg taatgacaag atacaatgat acatgcattc taagtaactc catgcgaaac
  481 gatcctaata ttgacttaca acatatagca ttttggctgt agtatattta ataatgttgt
  541 aggacacaac atgtagatta attgattaaa gacaacgtgc atgtctcttt tttttctcta
  601 tgatgcgggc cttttagcag aggaatgcta cagacatgta aagttttgat aaaattgaat
  661 aattagctaa ttatggcctg acttgtttaa cctctatatc cagaatgaac attatattga
  721 aaactgatct ttccacaact atatatgtaa tatagttttt tttgttttgt tgcattctga
  781 tctcgtcatt ctgatgtaat atagatcctt gccatcataa aaaaaaatgg taatcgattt
  841 ttagttaagt atgaagttgt cctcttagtt taaattacag tcagcgcaaa tgaattattc
  901 ttcccagaaa tcaattaatt ttcttgcatt ttaaaagcca gacataaagt tttgactgat
  961 taagacatac atgtaattta atttggttaa atgatttatc atttttctcg atcgattgga
 1021 taaaggacat ggttttctta gaaaagaata taagataaat gacaaataac agaaagaagc
 1081 gtgattgcct ttacttgcat gcagatgcag tagtctccca ttttgaatga ctgctctaaa
```

```
1141 ctcttaaagt ataactattc tcgtcaacta aagtttttgt ttcttttta acgtcaacta
1201 aagtctatat caataacaaa tatttatcgc ctaaccagta accatagata aagaaaaact
1261 tttaatctag aaaccattag cttctgaaac aaagtataaa cagcccttaa taaagccatg
1321 accatgtcaa ctgtttcgga ccatgtgtag tgtacgtagt aaacatttaa attgataatg
1381 aaatagcaag acgcaaatgg acgtttgtaa aattgtatgg ttggtaaaaa agatatgagg
1441 ctatatatga atttctgtca ttaaatcagc tttaggtatt acttgttagg attaaggtta
1501 tcgtagttga ttagactaat taactaatat atgaaccaaa aaagtataat agtcaggagt
1561 agtagttgag atttcaaaaa gattcaaggt ggccataaca caaatattat tcgcaatgac
1621 aagttgaagc atcgttttca cttttcaact tatccctctc aatccccacc accaccactc
1681 tctattccct catatagata gatcatccta gtgatcatat acacacacaa acattgatag
1741 agaacaaaac gaaattggaa aatttcattt taaattttaa aaatacaaaa ataaatggaa
1801 ttggaaaaag tctttttttt gtaaatacta atcaaatatt ttaaacttaa ccaaaaattc
1861 tatagtcaaa aaaagccttc tcaacatgat gtccacaatt cgaaaagtac atagtctttt
1921 tctcccttaa aaaatgatat gcaagctggc cggcgctact tgtgcgcaca cctacaagcc
1981 tatatataag atctctcact cgattcatgt aattatacac atatagagag attggagtga
2041 aaatattttt ttttaaatgg ttgtggtgaa caaaacgaat ctcttattgc ttcttctctc
2101 tctatgtctc accctcgacc tctcctccgc acaactccgc cgtaacttct acgccggaag
2161 ttgtcctaac gtcgaacaaa tcgttagaaa cgccgtacac aaaaaagtcc aacaaacttt
2221 caccaccatt cccgccactc tacgcctcta tttccatgat tgtttcgtca atgtaaaaca
2281 tatttttat tgatattatt ttcttgcgta aaattatcca ttatttggtc taaataattg
2341 attaatattc atcggggtgt gatgcatcag tgatgatagc gtcaacaaat aataataagg
2401 cggagaagga tcatgaggaa aatttatcgt tggccggaga tggattcgac accgtcatta
2461 aagctaaaga agctctcgat gccgtcccaa attgtcgtaa caaagtttca tgcgctgata
2521 ttctcacgat ggccactcgt gacgtcgtta atcttgttag taactcaagc ttatttattc
2581 tattaaatta aaacattagt ataagttatt aaattaaatt caatataggc gggtggacca
2641 caatacgacg tggaattggg aaggctagat ggactatcat caacagcggc tagcgtcgga
2701 gggaagttgc cgcatccgac cgacgatgtc aataaactca cttcactctt cgccaaaaac
2761 ggccttagcc ttaacgacat gatcgctctc tccggtaaga tttgaactac ttctcgatca
2821 taaatagtat cattcgaacc aaactaatag ctatgtgatt taatattccg gtttagctaa
2881 ctggatatcg gacccgaatt tgattagtat ttaagttggt ttatatatac ccgaatttga
2941 ttagtattta agttggttta tatttccggt cctaatttag atttcgtttg atcttaaaca
3001 tctcaggggc acacacatta ggattcgcgc attgcacaaa agtgttcaat cggatataca
3061 ctttcaacaa gacaactaaa gtggatccca cggttaacaa agattacgtg acagagctaa
3121 aagcgtcgtg tcctcgaaac atagatccaa gagtggctat taacatggac ccaacaacac
3181 caagacaatt cgacaacgtt tactacaaaa acttgcaaca agggaaagga ttgttcacgt
3241 cggatcaagt cttattcacg gatcgtcggt caaagccaac cgttgactta tgggctaata
3301 atggacagtt gtttaatcaa gcttttatta actcgatgat caagcttggt cgtgttggtg
3361 ttaaaactgg tagtaatggt aatatccgtc gtgattgtgg agcttttaat tgatttaaat
3421 atatctgttt tagattttgt ttattgtatt tgggtatagg atgggtcatg ggtctcataa
3481 agctggttca gctaaattga aaattgtaat aaatctctgt cttgggtttc agtttgcaaa
3541 acctaaactg taattataga gtggaggaaa taagaagcat cgtataatat ttatcaaggg
3601 atacatacag gccatccatt atgggccgcg taagagccaa tgcctaattt taataatttt
3661 cttaaacaaa ttattatgaa aaaggaccaa gaatgctgga tgcttcctac atacatataa
3721 agttttgata aaattaaaat aattaacgtc tatatccaga atgaagtcaa tgagcattat
3781 atatcccaaa ctgatctttc cacaactata caatatggat gtgttttttt tttcttttgg
3841 ttgcattctg atatgttcct tcttcttgtc tgcagcgatt acttgttatt ttgggttatt
3901 cgtggaaata caattcaacc atttgttttg aacgtgtaag ataagtaata tatggttaaa
3961 ttatttacca tttttctcaa tcgattagac aaaggacctg gtgttattag aatataatat
4021 aaatgacaaa caacagaaag aagcgtgatt gcccttactt agatttttt tggacaaacc
4081 catacttaga ttagatgcaa gagtagccct ttttgaatga ctgtaatctt gtcaactaaa
4141 gtttttgttt ctttttaaaa cgtcaactaa agtccatatc aatagctaat cgcctaatca
4201 tagataaaaa ctttaaaatc tatattaaac cattagcttc tgaaataaag taaaacagcc
4261 ctaaataaaa ccatgaccat gtcaactaat tcggacgacc atgtgtccat gtgtagcttt
4321 tcttttctag aactacgaga tccaaaatta agcagtgttt cataaggaaa cactaccgaa
4381 tgtaatgaat ttgatggata gatggtcacg ttatttttt gagataaagg tgtccagccg
4441 aaattaatac atataccaat tgttcaaaaa gtgacacatt tttgccatca aaatagacaa
4501 atgatgatga aatagcgcta gacgcaaatc ggagttttgt aaaaaaatgg tatggtaagt
4561 aacaagaata tgagagagag agactatatg aagtccttct tatcatcttt agttgtaact
4621 tgttaagatt gaggtttccg tagttgatta gactaattaa tattagttga tccaaaaaag
4681 tatgatattc aggagtagtg tagagatttc ggaaagattc aaggtgacca ttacacaata
4741 cttaaatact attcgcaatg acaagggaag catcgttatc aacttctccc tctcaatccc
4801 caccaccact ctctattacc tcatatagat agatcatcct agtgatcata cacacaacat
```

```
4861 tgatagcgaa caccaatagt atgctatatt cacaacaaaa caaaaatgga ctaattcatt
4921 tgtttttta agtaaaggaa ttgaacaata attattttat ttttttggg aaatactatt
4981 caagtatttg aaattagaaa agtatttagt ctttttctct ttaaaatatg aatgcaagct
5041 ggccgctggc cggcacacac agcacaaagc atgtatataa gatctctcac tccatttatg
5101 taattataca tataaagaga aattgattta tttttttaa tggttgtgat gaacaaaacg
5161 aatctcttgt tgcttactct ctcacttttt ctcgccatca acctctcctc cgctcaactc
5221 cgcggcgact tctacgccgg aacttgtccg aacgtcgagc aaatcgtcag aaacgccgta
5281 cagaaaaaaa tccaacaaac tttcaccacc attcccgcca ctctacgcct ctatttccat
5341 gattgtttcg tcaatgtaaa acaatctttt ctttgataaa aaatgatttc tagcgtaaac
5401 tatattttat aattttcggt aatttatttt attattatca tagggatgtg atgcatcagt
5461 gatgatagcg tcgacaaata ctaataaggc ggagaaggat catgaggata atttatcgtt
5521 ggccggagat ggattcgaca ccgtcattaa agctaaagaa gctgttgacg ccgtcccaaa
5581 ttgtcgtaac aaagtttctt gcgctgatat tctcacgatg gccactcgtg acgtcgttaa
5641 tcttgttagt aactccctcc atatttatat attataatgt tttattttaa tataaaaaca
5701 ttagtgtaag ttattacatt aaattcaaag taggcgggtg gaccacaata cgccgtagaa
5761 ttgggaaggc gagatggact ttcgtagtcg gcttctagcg tcacagggaa gctacccaaa
5821 ccgacctttg atctcaacca actcaatgca cttttcgcag aaaacggtct tagccctaat
5881 gacatgatcg ctctctccgg taagatgatt tgaactagtt cttgatcaag aatagtatca
5941 ttcggaccaa acaaattagc tatatgccat gttgttagcg gtttatattt ccggtggttt
6001 atatttccgg tcctaattta gatttcgttt gatcttaaac atctcagggg cacacacatt
6061 aggattcgcg cattgcacaa aagtgttcaa tcggatatac aatttcaaca agacaaataa
6121 agtggatccc acggttaaca aagattacgt gacagagcta aaagcgtcgt gtcctcgaaa
6181 catagatcca agagtggcta ttaacatgga tcccactaca ccaagacaat tcgacaacgt
6241 ttactacaaa aacttgcaac aagggaaagg attgttcacg tcggatcaag tcttattcac
6301 cgatagtcgg tcaaagccaa ccgttgactt atgggctaat aatggacagt tgtttaatca
6361 agcttttatt agctcgatga tcaagcttgg tcgtgttggt gttaaaactg gtagtaatgg
6421 taatattcgt cgtgattgcg gagctttcaa ttaatttaaa tatatctgtt taagatattg
6481 tttaattgca ctcatttgta ttgggatttg ggaatgggat gggtctcatg aggctggttc
6541 agctaaaatg taataattct ggttttgggt tttagtttgc tatatatcgc aaaactaaat
6601 tgtagatatg gagtggcaaa aaggaaataa tatacatgat tagtcccaat gggcaatgac
6661 aatactttgt atagtaagtt ggatcatgga gtcatgactt taatatttgt gtaaaatgtc
6721 aatcatctat tatttcataa gaatgtaagt agacataatt aatctttgtg ctgagaaata
6781 taacaaaatt acattttact tacaaaatta tcaaacgtat tgtcaaaaag aactaacaca
6841 aaatatagtt tctaattcat ataatatcaa cttccgagtt tcaacgtttt tgtatctttc
6901 ttgggatata acatataagg cccaaaccga ggcccatacc atagtaaagt aaggaaagca
6961 atggctttct tttgggcatg ttgtattttc cttataacaa tcttttgtag ctgctaggcc
7021 ctttggcaac aggccctttt tggaaacaat attagttttc gcttttgaaa aatgagaatt
7081 tgtcatgtgt aaccgtagtt tatccagaga ttatagtggc atactgatgc ctatttggct
7141 gcgtgaatcc aaaactaggc tataatgatt cacagagaca gagccacagt ggcataagca
7201 ttttaaactt aaatccaaaa tacatcaata aatcgtcaat caagttatgg tgttaccatt
7261 ttactagtag tggcttaact taaaaaactt taaggggcac acaagcccta gagattaatt
7321 gatccaacca agttcgactt caccgactgc ttttgacaac atttagaaca gctaaaagac
7381 actctttcca aatttggaaa atctaatatt tttgcttaaa actattttaa tgaattatca
7441 tcactaatga ttagtgacta cttcttttag ttgctaaact atattaatac gttagaatta
7501 tcataaagtt tgaaaaagta aaaaatatag caaaaaatga gcaaaacata aacaaaatat
7561 tatcaacaaa aataattata tattcaaatt tttagcaatg aaaaaaaaac attatgatta
7621 ttgattatca attacattaa aaattacacg aacacacggt gggagggatt cactgttcaa
7681 aaaatttttcc actgaatcta caagaaatca acggtcaata actcgacaat taacctgatg
7741 tttgcgacac cgaaaaagtt cgtctttatg gatttttca ggacttcgac aacgcatact
7801 atcgtaattt ggaagagtga aaaggagttt tggactccga tcaagttcta tacaccgccc
7861 cgcgctcacg accattagtt gaagaactgg ccagtaatgg tcaggttttc gaaagagttt
7921 tcgcacactc aatgattaag ttggacagaa tcggcgtgaa aacgggccgc aacagcaata
7981 tctgccttct ttgtgacgca tttaattagt tgtaacttgt gctgggcgtt gaaagcccat
8041 tctataatac tataatgtta ttaagacatg cagtaatgaa aatcttcatt taaaaaaaaa
8101 gaacctaaac tcttcacttc tcgtggtcca tgggttcgtt ttcgttgaaa gattttgatt
8161 cacgtatgac gtgaatggag tggtcgagag gttcccggtt ttttttttt attagatcgt
8221 gtaagttttt ttttgttta tgttgcaatt ttgagggtag tgttttattg cttagcctaa
8281 tatcacaaac gaagtcgttt taatagttta gaaagtatat aaacaaatcg atatcatctg
8341 aaacagtccc aatagaaaat taactttctt tgtagaataa tcaagatgtt tacgaactta
8401 ctaattgcaa acattcgcga ttaacttaca tcgttaggat ttgcatcaaa tgaagtaaag
8461 ttaccccaag attaatatta ttatcattat tatttaaaat acttaaattt ccacgatgca
8521 aatatatcga taacgttatt aaaaaaagta accgcatgat atattctctt tcgtatgata
```

```
8581 ttaaggctcc gaatggtgat tgcgttttgc gttacgggag aagtgcagtg cggttcaaat
8641 agtaacaaaa atacatagat acatatgtat atatacaaat tttaatactg ttacagtact
8701 attgcggtgc ggttttaaat gcggttttgg ttagcattca cagcctaatt actaaaggct
8761 ggagaaaaag atggtcatct tttgaaatcc attaattggc cacattttaa tattttaata
8821 tattatttta ataaaataaa taaatttatt atatattgag aaatacttta gaataacact
8881 gtacctttgg agacattcga taagctttag aataacatta aaataagtta tatggataat
8941 tatagcatac aatcaactaa cttttaaaat tagcattagt aaatttacaa aataaatatt
9001 tttttatcct atttaattaa aaaaataaaa ataaaaatta attaaatatt aaatctcaat
9061 tataatacga ctcatgaaat attaattcat ttgtgttatt ttaaaaaact ttatagctta
9121 tgtgataaaa gataacaaaa attatttttg tgtcatttgc gcgaattgcc aaaaaaaaat
9181 aggtttgact gatgcaaaaa aatggctatt tgcattatcc aaaagccgtc ttatccatga
9241 aaaccaaaat ttcggtgtat tattcatgta aaaattactg tattgatata aaatattggc
9301 aaatatcgta ctaatgtcac ctggaaattt ggaatccctc ggaagctaat ttatcgattt
9361 gcttcttcac aaatcaagtg tcaactcttt atatgatttt attcgtattt aggatcttta
9421 tatttcatgt gatctgctca gtgggaaaaa ataaaaacaa gtacaaaaca cacggttgga
9481 actgtagagc ggggatgtta caggagctat aggcattttc cttcattctc aatttctaga
9541 tttttcttac tgcattagga tggtggtagt agctaactta tggaaattat gtgatatttt
9601 cagttattgt cgcgttacta tcgttttact ctattagtat gaatcagata aggcctaatt
9661 agttttctac acacagagtc cattagtggc tgaatgttga tccaattttc acgaatatca
9721 tatatgatca gttcttaata ttgtagaaga tacactatgt gtgatatgtc ctctttaaat
9781 gagtttccct gataatttat aatatttgaa tagttaagct tggagaaaaa aacatacaca
9841 ccatatgtta gcaatgaaac aattctcaac agattattta ttctcatcgg caaaatatat
9901 atccccaaaa ttttatcata tttctcgata gagaaaaatc attaactaaa tatccattcc
9961 cactcaaaac cctaaaaaag agagcctcaa aatctttgag tgttttgcgt cgtcggtttt
10021 aaacaaaaag gttcaataac tttttctctc ctttgtttct tttttgttta ttatctcctc
10081 tcatcccatt ttcccaacac gacgtcgtta cagattcctc cgatggactt cccggacgat
10141 caccccttaac caccttacaa aacgacgacg gaggaagaaa aggaaccggt acagccggag
10201 cggttagggt taaaccgtgg ttcacttgga gatcaagatc cgaatcggac ttcaatcgtt
10261 ttcggttaac cagagagtct ttcatctccg tcgtggatct ggaggtggag aagcgacaat
10321 ggtgggtttg gtttcttgaa acgtcttgat gccaaatctc tgaagactca tcacaggtag
10381 atcatcaacg acggagattc aagaaagatc tgatatcggt cgtaaagttc cgccgcgaaa
10441 gaacagtctc aaccgccgct tggcattcgt gccaagttcc ctggtccaca acattccaac
10501 cgctccgtta accggattca ccgttctccc acacacgctt caaaacaaac aacgactgaa
10561 acaaagctga acaaaccaaa acataaaaat caatttaaaa tcaaaaacat taaaagttgt
10621 atagatccgt agagtttgta acctaatggt tatgttatat attcgaataa aaatagatag
10681 tatagtgata aaaagaagtt gtaccaggac gttggagttc aggtacggag gagatgaaag
10741 acatgagacc agcacgacca aagaatttag cgacgaagac ggtggcgtgg ccttgtgact
10801 cggcggattc gatccattga aggcaaggac gaaggatgca tgtttcactg caacctttc
10861 gaagaactct acatccattg caactcattt cttttccgaa atttgaaaaa gaaatatatc
10921 caatattttg ggaaactcta acttgttaaa acaaaaactg gaaaagtgtc gatttaattc
10981 gggtttgaga gcatgcaaat aggttttgat tgaagaaggg aagagaaaac aagtttcaag
11041 aagatgagag aagatagtat gagatttata gaggggagtg aataataaag agagtgtgcc
11101 aacttgtgtt tctttttat tcctttgtcc tttcttttcc attatttcta attaaaaaac
11161 catgtgtatt aattatttga aatttattat ggggaagaca gaaactcaaa ggagattgag
11221 ataacatttt gaaaagttaa aatacttatg tgaataaaga atgggtcata ttgacacaaa
11281 taaactagtt tttcgatcat taattcaaac ttgagttgat ttataagttg ctggtttcaa
11341 ttttattggt gtgtagaaaa tgtaaaaaga gagaaagtga ggacgaaagg acaggacaga
11401 gttagggagg gattccaaag gtattggcgt tgggggaaga gagagaaaga gagcgtttcc
11461 gtttctccgc gaacacttgt cgctgactca gactcctctt atattagtat tactttcttc
11521 ttttaagtta aagtattaaa gatgtgtact cttctctttt ttataagagt acgta
//
```